Journal of Development Economics

Building Teacher Capacity or Outsourcing Pedagogy? Experimental Evidence from Science Education Reform

--Manuscript Draft--

Manuscript Number:	DEVEC-D-25-02418R1
Article Type:	Registered Report Stage 1: Proposal
Section/Category:	Health, Education, gender, poverty
Keywords:	Human Capital; Curiosity; Critical thinking; Teacher training; External instructors; education production function; India
Corresponding Author:	Nishith Prakash, Ph.D., Northeastern University Boston, MA UNITED STATES
First Author:	Nishith Prakash, Ph.D.,
Order of Authors:	Nishith Prakash, Ph.D.,
	Santosh Kumar Gautam
	Ariel Gomez, PhD
	Sandeep Kumar
	Soham Sahoo, PhD
	Vinay Kumar Jha, PhD
Abstract:	Skills such as curiosity and critical thinking empower students to explore ideas and adapt to change, yet fostering these skills in low-capacity schools remains challenging. We evaluate two delivery models for curiosity-based science education through a randomized trial spanning 150 government middle schools and 7,185 students in rural Uttar Pradesh, India. We compare teacher training—where science teachers receive intensive workshops and ongoing support to integrate inquiry-driven pedagogy—against external instruction, where specialized facilitators deliver the science curriculum directly to students. Both interventions use identical content aligned with the school syllabus and grounded in cognitive science research on curiosity. The design addresses a central policy question: should reforms build internal teacher capacity or rely on external agents? We measure impacts on curiosity, critical thinking, growth mindset, and academic performance. By tracking classroom practices, we situate this comparison within the realities of public education systems characterized by workload constraints, hierarchical structures, and limited instructional support, providing evidence for scalable approaches to strengthening pedagogy in low-capacity settings.
Response to Reviewers:	

Building Teacher Capacity or *Outsourcing Pedagogy*? Experimental Evidence from Science Education Reform

January 14, 2026

Abstract

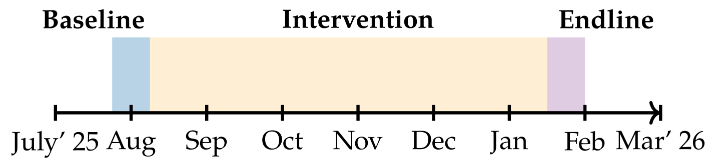
Skills such as curiosity and critical thinking empower students to explore ideas and adapt to change, yet fostering these skills in low-capacity schools remains challenging. We evaluate two delivery models for curiosity-based science education through a randomized trial spanning 150 government middle schools and 7,185 students in rural Uttar Pradesh, India. We compare teacher training—where science teachers receive intensive workshops and ongoing support to integrate inquiry-driven pedagogy—against external instruction, where specialized facilitators deliver the science curriculum directly to students. Both interventions use identical content aligned with the school syllabus and grounded in cognitive science research on curiosity. The design addresses a central policy question: should reforms build internal teacher capacity or rely on external agents? We measure impacts on curiosity, critical thinking, growth mindset, and academic performance. By tracking classroom practices, we situate this comparison within the realities of public education systems characterized by workload constraints, hierarchical structures, and limited instructional support, providing evidence for scalable approaches to strengthening pedagogy in low-capacity settings.

Keywords: Human capital, Curiosity, Critical thinking, Teacher training, External instructors, Education production function, India

JEL Codes: I21, I25, I28, J24, O15

Study pre-registration: AEA RCT Registry RCT ID: AEARCTR-0017098.

Proposed timeline (*required*):



1 Introduction

Despite widespread recognition that 21^{st} century skills—including critical thinking, curiosity, communication, collaboration, and problem solving—are essential components of human capital (Heckman and Kautz, 2012; Deming, 2017; Jackson, 2018), effective and scalable strategies to teach these skills in low-capacity public schools, particularly in low- and middle-income countries (LMICs), remain unclear. Global policy frameworks, ranging from the United Nations' Sustainable Development Goal 4 to India's National Education Policy (NEP) 2020 (Ministry of Education, 2020), explicitly call for cultivating these higher-order abilities among students. Yet classroom pedagogy in most LMICs remains misaligned with this goal, relying on teacher-centered, exam-oriented instruction that rewards rote memorization over inquiry, experimentation, or classroom discussions. The real challenge for policymakers is not in recognizing the value of these skills, but in finding practical and proven ways to teach them in under-resourced public schools, where teachers already face heavy administrative workloads, large classes, and little instructional support.

To address these institutional and pedagogical constraints, this paper evaluates two scalable delivery models designed to foster 21^{st}-century skills among middle-school students (grades 6-8) through a randomized controlled trial in Uttar Pradesh (UP), one of India's largest and most populous states, where learning levels in government schools remain persistently low. Developed in partnership with the Agastya International Foundation (AIF), a leading NGO in experiential science education,[1] and implemented by the EcoPrism Collective Foundation (Ecoprism)[2], the intervention delivers an identical curiosity-driven science curriculum through two pedagogical models that differ only in their mode of delivery. Designed by AIF, the curriculum comprises ten hands-on, inquiry-based lessons explicitly mapped to the UP state's syllabus for grades 6–8. Each lesson is structured to promote observation, questioning, experimentation, and reflection—encouraging students to connect scientific concepts with their everyday experiences and to think critically about evidence and explanation. The curriculum design draws on cognitive-science research on curiosity, metacognition, and inquiry-based learning (Perkins, 1992; Swartz et al., 2017), providing a structured framework for cultivating

[1] Agastya International Foundation (AIF) is a nonprofit dedicated to experiential learning in India, operating mobile and school-based programs that emphasize learning through doing and discovery. By 2032, it aims to equip 100 million children and 1 million teachers with creativity, critical thinking, and innovation skills, particularly in underserved communities.

[2] Ecoprism Foundation is a nonprofit organization based in Varanasi, Uttar Pradesh, dedicated to advancing sustainable education and community development. Its programs promote environmental awareness, climate literacy, and youth engagement, particularly in underserved regions of India.

21st-century skills within regular classroom settings.

Both intervention arms deliver identical content but differ in pedagogical delivery: (1) **Teacher Training (T1)**—regular government science teachers receive four days of intensive training and ongoing support to integrate and facilitate the curiosity-based pedagogy developed by AIF within their existing instruction; and (2) **External Instructor (T2)**—external facilitators trained by AIF to deliver the same lessons directly to students while regular teachers continue with business-as-usual instruction. This design isolates the trade-off between investing in teacher capacity—which can increase workload and cognitive burden—and circumventing bandwidth constraints by outsourcing delivery to specialized external agents. It enables us to address a central policy question: *Should education reform efforts strengthen the capacity of existing teachers through pedagogical training, or rely on external agents who can deliver specialized instruction?* More broadly, in systems characterized by organizational frictions—such as workload constraints, professional hierarchies, and perceived threats to teacher competence—how do these alternative delivery models shape the effectiveness of pedagogical reforms in resource-constrained environments? Consistent with evidence from Ganimian et al. (2025), who document complementarities between teacher content knowledge and pedagogical supports in India, our design contrasts models that either build capacity among existing teachers or rely on specialized external instructors to deliver high-quality instruction.

We conceptualize our intervention within the framework of educational production functions that link school, teacher, and student inputs to the formation of cognitive and non-cognitive skills—broadly encompassing curiosity, critical thinking, communication, collaboration, and problem solving (Hanushek, 2020; Glewwe et al., 2020). Within this framework, curiosity, defined as the intrinsic motivation to acquire new knowledge in response to information gaps (Keller et al., 2019; Lins de Holanda Coelho et al., 2020; Alan and Mumcu, 2024), serves as an endogenous input that enhances students' effective learning effort and persistence, thereby increasing the marginal productivity of instructional time and materials. The proposed course content and pedagogical strategy seek to shift the underlying production function outward by inducing deeper cognitive engagement and transforming classrooms from spaces of passive memorization to environments that foster inquiry and curiosity.

The intervention in our study draws on Self-Determination Theory (Ryan and Deci, 2000), which posits that autonomy, competence, and relatedness drive intrinsic motivation. Our intervention integrates group-based experiments, guided discovery discussions, and student-led demonstrations—pedagogical features that foster both autonomy and collaboration. In economic terms, these features relax the constraints on the *quality* of teaching (through new pedagogical tools) and the *returns to effort* for students (through increased intrinsic motivation), thereby improving the productivity of both inputs. As motivation strengthens, learning effort can increase inside and outside the classroom, potentially crowding complementary behaviors such as peer study and more time spent on science activities.

Teacher quality is the key input in the educational production function, yet institutional conditions in many LMICs constrain it (Hanushek, 2020; Glewwe et al., 2020). In India, poor management, limited autonomy, weak incentives, and low accountability reduce instructional quality; absence and reduced instructional time are common, with substantial time diverted to non-teaching tasks (Chaudhury et al., 2006; Kremer et al., 2013; Muralidharan et al., 2019; Abadzi, 2009; Bruns and Luque, 2014). Beyond these factors, another key bottleneck is inadequate teacher training and content mastery. Gaps in pre-service training and sporadic in-service workshops leave teachers with limited content mastery and pedagogical content knowledge, little practice with inquiry-based teaching methods, and minimal classroom coaching. Without adequate training in curiosity-based pedagogical practices, even motivated teachers struggle to plan lessons, manage active learning, and adapt instruction, resulting in lower instructional quality and limited acceptance of new pedagogy.[3] Consequently, pedagogical reforms often fail to translate into sustained practice (Yoon et al., 2007; Loyalka et al., 2019; Fryer, 2017). These challenges motivate our comparison of two delivery models: one that targets the quality of current teachers against another that bypasses teacher capacity by having external facilitators deliver the curriculum.

In the *teacher-training arm*, incumbent grades 6-8 science teachers attend structured training programs developed by AIF and implemented by Ecoprism that model and cultivate curiosity-based pedagogy. The training program strengthens content mastery, introduces active-learning techniques, and provides guided practice. We hypothesize that as students respond to the new pedagogy, teachers will update their beliefs about their own efficacy and the returns to effort. These beliefs will in turn shift teachers toward more student-centered instruction—a mechanism consistent with models in which self-efficacy enhances teacher productivity (Glewwe et al., 2020). Ongoing peer support (moderated

[3] Our baseline survey indicates that status-quo pedagogy relies heavily on memorization, operates with weak instructional support, and provides minimal time for lesson preparation.

WhatsApp groups) and ready-to-use materials reduce the cost of sustained adoption, allowing spillovers as teachers internalize methods, share practices, and accumulate pedagogical capital within schools. In the *external-instructor arm*, AIF-trained facilitators deliver the same curriculum in ten structured sessions. This model circumvents teacher capacity and incentive constraints, ensuring greater short-term fidelity and standardization. From a policy perspective, this represents contracted delivery—analogous to the hiring of specialized agents to provide high-quality input into the education process. Although external provision may achieve higher immediate learning gains, its sustainability depends on whether the curiosity and collaborative behaviors it induces persist after facilitators exit and whether regular teachers later adopt similar practices.

Yet, introducing external instructors into government classrooms is not without organizational friction. Research in organizational psychology and public-sector management shows that external interventions can crowd out intrinsic motivation when workers perceive threats to autonomy or competence (Frey and Jegen, 2001). Regular teachers may view NGO facilitators as a critique of their skills or a challenge to their authority, reducing engagement or cooperation. Threats to professional status and autonomy can erode commitment and increase emotional exhaustion (Firestone and Pennell, 1993; Skaalvik and Skaalvik, 2017). Evidence from contract-teacher programs in India and Kenya reveals tension and reduced cooperation between regular and externally hired teachers (Muralidharan and Sundararaman, 2015; Duflo et al., 2015). The organizational context matters: Bold et al. (2018) show that NGO-managed contract teachers outperform government-managed ones, underscoring how internal frictions shape outcomes. In the broader public sector, Rasul et al. (2021) find that autonomy and task clarity predict performance in Ghana's civil service, while theoretical work highlights that performance comparisons with external agents can dampen mission-driven motivation (Delfgaauw and Dur, 2008). Together, this literature suggests that externally delivered pedagogy may relieve capacity constraints but can also displace teacher ownership over learning, creating symbolic crowd-out and morale costs. In the language of motivation crowding theory (Frey and Jegen, 2001), external provision risks violating autonomy, competence, and relatedness—the psychological foundations of intrinsic motivation. The resulting equilibrium response may offset short-term efficiency gains by weakening cooperation and long-term sustainability.

The two delivery models thus embody a policy-relevant trade-off between *building internal capacity* and *augmenting delivery through external expertise*. Teacher training represents a long-term investment in public-sector capacity, but faces challenges of cost, time, and implementation quality. External facilitation can ensure more consistent short-term delivery and demonstrate effective pedagogical practice, yet may have limited persistence and weaker institutional integration—a central tension in debates on service delivery in low-capacity systems (Muralidharan and Sundararaman, 2013; Glewwe et al., 2020; Hanushek, 2020; Romero et al., 2020). Both approaches could generate broader system-level effects. Training may lead to peer or inter-cohort spillovers, while external facilitation could influence student engagement and parental perceptions of science learning. By comparing these two models, the experiment directly tests how alternative delivery routes for experiential pedagogy affect curiosity, engagement, and learning outcomes, thereby informing the broader debate on how best to strengthen human capital formation in resource-constrained settings (Duflo et al., 2015; Bold et al., 2018).

A growing body of evidence underscores these patterns. Curiosity enhances attention, retention, and the transfer of knowledge (Loewenstein, 1994; Keller et al., 2019). Interventions that foster curiosity and related socio-emotional skills show promising results: Alan and Mumcu (2024) find gains in science achievement and willingness to pay for knowledge in Turkey; Bharti et al. (2024) show that mobile science laboratories in India increase confidence and interest in science; and Nourani et al. (2025) find that inquiry-based curricula in Uganda improve both reasoning and test scores. Yet these studies also highlight substantial implementation demands—success depends on sustained teacher engagement and structured support. Parallel evidence from teacher-incentive programs reinforces this view: while contract teacher hiring or external facilitation can improve learning outcomes (Muralidharan and Sundararaman, 2013; Duflo et al., 2015; Bold et al., 2018), effectiveness hinges on organizational design and frontline cooperation. Public-management research echoes these insights, showing that overburdened bureaucracies reduce implementation quality when staff face competing directives (Giauque et al., 2013; Ahmad et al., 2025).

This study directly addresses that gap. Beyond evaluating learning outcomes, our design also measures changes in curiosity, critical thinking, and growth mindset—non-cognitive traits increasingly recognized as valuable components of human capital—and examines how alternative pedagogical approaches influence their development (Alan and Ertac, 2018; Alan et al., 2019; Sorrenti et al., 2024). By combining experimental variation in the content and structure of the pedagogical delivery with novel psychological measures, we make four contributions. First, we provide rare experimental evidence that compares teacher training and external delivery within an identical intervention—central to policy debates on scaling effective models. Second, we integrate curiosity and metacognitive processes into the economics of human capital production, expanding the framework of skill formation beyond test scores. Third, we identify organizational frictions—including motivation crowd-out and cooperation failures—as mediating factors shaping reform success. Fourth, we offer practical insights into how teacher workload and pedagogical complexity interact, informing scalable policy design in low-capacity systems.

Ultimately, our objective goes beyond testing whether curiosity can be taught. We aim to understand how curiosity-based pedagogy can be institutionalized and sustained in systems where teachers are already overburdened. By experimentally comparing teacher training and externally delivered instruction under identical curricular content, we identify feasible pathways to embed 21st-century skills within the organizational realities—and frictions—of public education in the developing world.

2 Background

We conducted this study in Fatehpur district, located in the state of UP, India.[4] Fatehpur is predominantly rural, with 87.8% of its 2.63 million residents living in rural areas—higher than the state average of 77.7% (Census, 2011). In terms of core socio-educational characteristics, however, Fatehpur closely resembles a typical rural district in UP. The district's overall literacy rate is 67.4%, nearly identical to the state average of 67.7%, with comparable gender gaps (77.2% for males and 56.6% for females).

Consistent with this demographic profile, educational participation and learning outcomes in Fatehpur closely track statewide rural averages. According to the ASER 2024 Rural District Estimates, 53.5% of children aged 6–14 in Fatehpur are enrolled in government schools, compared to a rural UP average of 49.1%, indicating similar reliance on the public school system. Learning outcomes are likewise comparable to state norms. Among children in grades 3-5, 43.1% in Fatehpur can read a grade II–level text (state average: 46.5%) and 50.3% can perform at least subtraction (state average: 52.7%). Among children in grades 6-8, 65.1% can read a grade 2–level text (state average: 68.8%) and 44.6% can perform division (state average: 49.4%) (ASER Centre, 2024). Taken together, these

[4] Uttar Pradesh is India's most populous state, with over 240 million residents, and has one of the highest poverty rates in the country. Despite recent progress, it continues to face major development challenges, including widespread undernutrition and limited access to quality health services, particularly in rural areas.

indicators place Fatehpur close to the center of the distribution of rural districts in UP in terms of foundational learning.

Within Fatehpur, the intervention was implemented in four rural blocks—Bahua, Bhitaura, Haswan, and Telyani—which together account for 27.3% of the district's population. Literacy rates in these blocks (67.1%, 64.8%, 63.2%, and 68.3%, respectively) mirror both district-level and statewide rural patterns, indicating that the study sample reflects typical within-district variation rather than unusually high- or low-performing areas.

Despite near-universal enrollment, learning outcomes in UP remain persistently low. ASER 2024 reports that only 34% of grade 3 students in government schools in rural UP can read a grade 2 text, and only 31% can perform a two-digit subtraction task. By grade 5, just 56% can read a grade 2 text and 39% can complete a division problem—figures that, despite modest improvements since 2022, underscore a substantial deficit in foundational learning. Private-school students perform better (66% and 52%, respectively), but much of this gap reflects socioeconomic selection rather than intrinsic differences in school quality (Kumar et al., 2024). School-level constraints remain widespread: in 2024, only 63% of primary schools in UP had usable libraries, 89% had usable toilets, and 77% had access to electricity (ASER Centre, 2024).

These weaknesses are deeply embedded in the governance and pedagogical structures of public education in UP. Kingdon and Muzammil (2009) describe the state's schooling system as characterized by entrenched teacher-union influence, weak local accountability, and highly centralized management—features that have persisted despite successive waves of reform. Administrative oversight has often emphasized procedural compliance rather than instructional quality, while classroom practices remain heavily textbook-driven and exam-oriented, leaving limited scope for inquiry, experimentation, or creative problem-solving.

Recognizing these constraints, the NEP 2020 explicitly calls for a shift toward experiential learning, critical thinking, and the cultivation of scientific temperament (Ministry of Education, 2020). However, translating these principles into routine classroom practice has remained uneven, particularly in rural and resource-constrained settings. Evidence from neighboring districts illustrates both the depth of these challenges and the scope for improvement. For example, Kumar et al. (2024) show that strengthening community–school collaboration in Sitapur district led to substantial gains in foundational literacy and numeracy, with especially large improvements among students from disadvantaged backgrounds. These findings underscore the role of complementary institutional and behavioral interventions in overcoming persistent system-level constraints.

Beyond governance and community engagement, a growing literature emphasizes the limited capacity of teachers to adapt and update their pedagogical practices. As Muralidharan (2024) argue, public-sector teachers typically remain in service for several decades after recruitment, making continuous in-service professional development central to efforts to strengthen state capacity. Yet, in most LMICs settings, such training remains sporadic, poorly designed, and weakly linked to classroom realities. Synthesizing evidence across countries, Glewwe and Muralidharan (2016) emphasize that sustained learning gains arise not from additional inputs alone, but from improvements in the quality of instruction and classroom interactions.

Our training program is designed with these insights in mind. It emphasizes hands-on experimentation, guided inquiry, and peer-based learning to strengthen teachers' ability to implement experiential, curiosity-driven science instruction aligned with the NEP. By embedding these practices within a structured, curriculum-aligned model that operates inside the government school system, the intervention aims to build durable pedagogical capacity while confronting the institutional constraints that characterize not only Fatehpur, but many rural districts across Uttar Pradesh.

3 Intervention, Sample Selection, and Experimental Design

3.1 The Intervention: Curiosity Pedagogy

The intervention is grounded in a curiosity-based science pedagogy jointly developed by the research team and the Agastya International Foundation, an Indian NGO known for its pioneering work in experiential and inquiry-driven science education, and implemented by the EcoPrism Collective Foundation (see Appendix A.4 for detailed session descriptions). Agastya has a long history of designing scalable, low-cost science modules for government schools in India that promote scientific reasoning through discovery, experimentation, and reflection. For this study, Agastya collaborated with the research team to adapt its existing modules to the UP state curriculum for grades 6–8 (equivalent to middle school in the United States). The resulting curriculum consists of ten carefully sequenced lesson plans that together constitute a structured program of scientific inquiry. Each session follows a common pedagogical arc: beginning with a "super start"—a striking demonstration or provocative question that triggers curiosity—followed by guided hands-on experiments, collaborative reasoning, and reflective discussion. Through this sequence, the intervention shifts classroom practice from rote memorization and recitation toward active learning and student-led exploration.

The curriculum spans foundational topics in physics, chemistry, and environmental science that are integral to the upper-primary syllabus. The early sessions introduce tangible and observable phenomena drawn from the immediate environments of the students. For instance, lessons on soil composition and water absorption prompt students to collect local soil samples, observe texture and sedimentation, and measure how different soils retain water – linking textbook content directly to local agricultural experience. Subsequent modules on acids and bases encourage students to construct natural indicators using turmeric, testing household substances such as lemon juice, soap, and vinegar to identify their properties. These activities connect abstract chemical principles to everyday life, demonstrating that science is embedded in daily routines and accessible through observation and experimentation.

As the sequence progresses, students engage with topics that require progressively higher levels of abstraction and reasoning. Experiments on heat transfer, using wax, colored water, and simple metallic rods, help students visualize conduction and convection, while sessions on air and water pressure allow them to infer invisible forces through tactile and visual cues, such as inverted tumblers or perforated bottles. A set of modules on magnetism introduces the concept of invisible fields through the creation of temporary magnets and explores factors affecting magnetic strength. Later sessions transition toward the microscopic and theoretical: students model the atomic structure using simple materials such as colored buttons, understanding the relative size of protons, neutrons, and electrons, and relating these to the organization of the periodic table. The final lessons emphasize the conservation and distribution of natural resources, including the water cycle, filtration, and sustainable use, integrating environmental awareness with scientific reasoning.

Throughout all sessions, the pedagogical design reflects a deliberate emphasis on *learning by doing* and *thinking by questioning*. Each activity invites students to hypothesize, observe, record, and explain phenomena in their own words. Facilitators or teachers are trained to guide rather than instruct—encouraging students to predict outcomes, discuss results in small groups, and draw conclusions collectively. The design also embeds opportunities for guided reflection, where students relate classroom findings to familiar contexts—why soil texture matters for farming, how heat moves through cooking utensils, or why magnets weaken when heated. This continuous movement between concrete experience and conceptual understanding is intended to foster cognitive flexibility and sustained curiosity. By promoting dialogue, hands-on exploration, and iterative reasoning, the pedagogy nurtures both cognitive and non-cognitive dimensions of scientific learning, including perseverance, creativity, and a sense of ownership over knowledge.

Importantly, *Curiosity Pedagogy* was developed to function effectively in resource-constrained public schools, where laboratory equipment, teaching aids, and time for experimentation are typically scarce. Most materials used in the sessions are locally available and inexpensive – plastic bottles, nails, candles, turmeric, and soil samples – making the approach feasible for scale within existing school budgets. Each activity is designed to be conducted within a single classroom period and to require minimal preparation time, allowing it to fit smoothly into the government schedule. Beyond individual lessons, the intervention seeks to shift classroom norms: repositioning the teacher as a facilitator of inquiry rather than a transmitter of facts, and empowering students, especially girls and those from disadvantaged backgrounds, to participate actively, ask questions, and express reasoning publicly. In this sense, *Curiosity Pedagogy* is not only a pedagogical tool but also an institutional innovation aimed at democratizing access to scientific thinking and restoring a sense of wonder and exploration to everyday learning.

3.2 Sample Selection

The study sample comprises 160 science teachers and 7,185 students enrolled in grades 6-8 across 150 government schools in Fatehpur district. These schools are located in four centrally located blocks of the district—Hasawa, Bahua, Bhitaura, and Teliyani. Figure 1 shows the location of Fatehpur within the Indian state of Uttar Pradesh.

Figure 1: Location of the Fatehpur District in the State of Uttar Pradesh

3.3 Randomization

The study follows a cluster-randomized design, with randomization conducted at the school level and stratified by administrative block. A total of 150 schools were randomly assigned in equal numbers to one of three groups: 50 schools in Treatment 1 (T1), 50 schools in Treatment 2 (T2), and 50 schools in the control group (C).[5]

In the 100 treatment schools, the intervention will be implemented during the 2025–2026 academic year, from August 2025 to January 2026. The 50 control schools will continue with standard classroom instruction and will not receive any component of the intervention. Figure 2 depicts the spatial distribution of treatment and control schools across Fatehpur district.

[5] Randomization was implemented using the `randtreat` command in `Stata`, stratified at the block level. The algorithm prioritized global balance across treatment conditions rather than exact balance within each block. In practice, the number of schools per group within each block differed by no more than one.

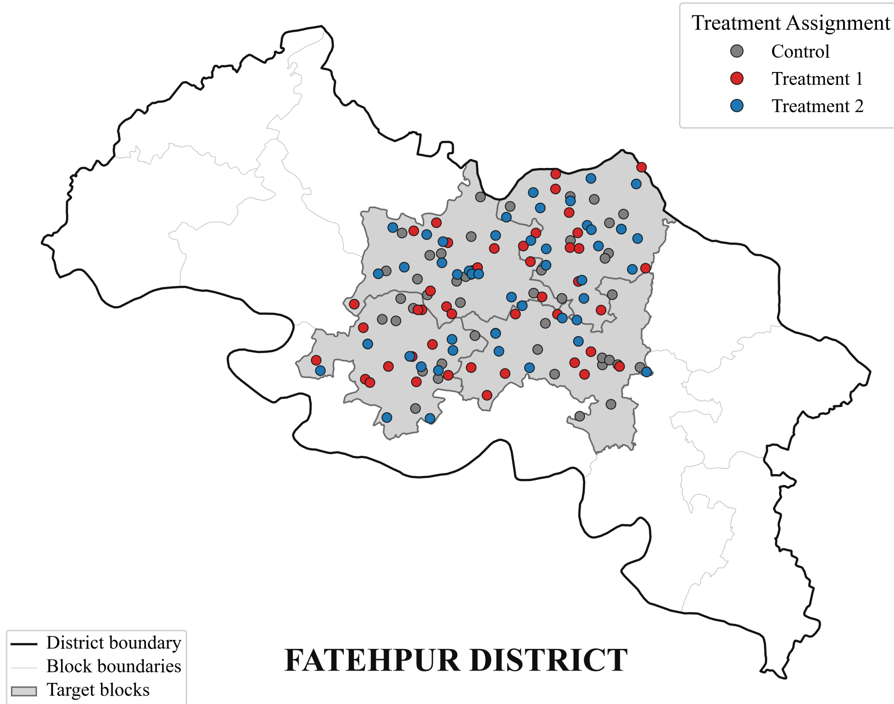


Figure 2: Geo-Spatial Distribution of Treatment and Control Schools in the District of Fatehpur

3.4 Treatment Groups

Rural public-school teachers in India, including in our setting, face substantial constraints: large class sizes, limited training opportunities, and inadequate resources, which make it challenging to independently adopt and sustain new pedagogical approaches. Accordingly, our experimental design compares two models of pedagogical delivery that mirror the implementation choices policymakers typically face when scaling pedagogical reforms: whether to invest in building capacity among incumbent teachers or to outsource instructional delivery to specialized external instructors. The intervention is designed to test alternative models for implementing a *curiosity-based science pedagogy* that integrates activity-based learning with the existing state curriculum and aligns with the objectives of India's 2020 National Education Policy. In both treatment arms, students are exposed to the curiosity-based pedagogy repeatedly over a sustained five-month period (September 2025–January 2026), integrated into regular school instruction. The core contrast between the two arms lies in who delivers the pedagogy and how it is embedded within the school system, rather than in the length of exposure itself.

Treatment 1 (Teacher-Training Model): In the first treatment arm (T1), science teachers were invited to participate in a structured, in-person training program aligned with the state science curriculum and based on curiosity-driven modules jointly developed by the research team and the Agastya International Foundation (AIF). The AIF team first trained a group of science educators from the Ecoprism Foundation (hereafter "facilitators"), who then conducted the teacher training sessions to ensure pedagogical fidelity and continuity during field implementation. Training inputs were delivered in multiple phases over the intervention period and were explicitly timed to coincide with classroom implementation. The multi-day sessions were designed to strengthen teachers' conceptual and pedagogical understanding of the curriculum, enhance their ability to foster inquiry in the classroom, and familiarize them with hands-on, activity-based learning methods.

During the training, teachers actively engaged with experiments from the curriculum, discussed context-specific adaptations for diverse classroom settings, and reflected on strategies to cultivate curiosity-driven dialogue among students. Following this, teachers participated in a Curiosity Pedagogy Workshop that formally introduced a curiosity-inducing framework emphasizing inquiry, observation, and collaborative experimentation. Teachers received a month-wise activity schedule aligned with the science curriculum and performed selected experiments in small groups to simulate classroom implementation. Each teacher was also provided with a detailed activity handbook and low-cost materials designed for repeated use across grades 6–8 throughout the intervention period. Teachers were expected to implement these activities regularly in their own classrooms over the five-month window, rather than replicate a single demonstration or lesson.

To promote sustained professional development, teachers were added to a dedicated WhatsApp group that served as a professional support network where they could share activity photographs, clarify doubts, exchange ideas, and seek real-time guidance from facilitators. This digital channel supported ongoing engagement with the pedagogy, reinforced training inputs over time, and helped create a collaborative community of practice among participating teachers. Recognizing that teacher training alone cannot succeed without institutional support, a Headmaster Workshop was also held for all schools in T1. These sessions introduced the project's objectives, demonstrated sample classroom activities, and emphasized the role of school leadership in sustaining curiosity-based learning. Headmasters were trained to monitor activity schedules, encourage teacher participation, and facilitate coordination during the intervention period. All participating headmasters were added to a separate WhatsApp group to receive regular updates, coordinate activity timetables, and share experiences related to implementation challenges.

Treatment 2 (External-Instructor Model): In the second treatment arm (T2), facilitators belonging to the same group of trainers who led the teacher training in T1 directly visited schools to conduct science sessions with students. This design ensured that the content, pedagogy, and materials were identical across both treatment arms, allowing for a clean comparison of delivery models. The key difference lies in the agent of delivery: in T1, incumbent teachers implemented the activities themselves after receiving training and ongoing support, whereas in T2, trained facilitators delivered the same activities directly to students.[6] Regular classroom instruction otherwise continued as usual, and teachers in T2 schools did not receive formal training or structured exposure to the curiosity pedagogy. As a result, facilitator-led sessions in T2 augmented rather than replaced regular instruction, and total instructional time could exceed that in T1.

Facilitators conducted curiosity-based lessons following the structured modules over the same five-month period as T1, with sessions spread over time rather than concentrated in a short burst. These lessons were delivered separately for each grade, and because the intervention targeted grades 6-8, each T2 school received multiple facilitator-led sessions per grade, resulting in repeated exposure to the pedagogy across the intervention period. Facilitators returned to the same schools multiple times, allowing for sustained engagement with students and consistent implementation of the curriculum. Throughout the intervention, the implementation team remained in regular contact with facilitators, headmasters, and schools to monitor scheduling, attendance, and fidelity of delivery. We assess whether regular teachers adjusted attendance or instructional practices in response to facilitator presence using unannounced spot checks and classroom observations during regular classes.

A Headmaster Workshop similar in structure to that in T1 was also conducted for T2 schools. These sessions introduced the program objectives, demonstrated key classroom experiments to familiarize headmasters with the nature of facilitator-led activities, and outlined their coordination and monitoring roles. In T2, this orientation did not signal any expectation that regular teachers would adopt or implement the curriculum; rather, it was intended solely to enable external instructors to conduct the sessions as planned. As in T1, participating headmasters were added to a WhatsApp coordination group to share implementation updates, address scheduling issues, and provide feedback related to facilitator visits and session logistics.

[6] The same curriculum-aligned instructional materials used in the teacher-training arm are also used by external instructors in T2 when conducting the corresponding science experiments.

Control Group: The remaining 50 schools formed the control group and continued with business-as-usual classroom instruction. Teachers in these schools did not receive any form of training, materials, or exposure to curiosity pedagogy. The control group thus provides a benchmark for evaluating the marginal effects of the teacher-led and facilitator-led models of implementation on both student outcomes and teacher engagement.

3.5 Incentives and Contractual Structure

Neither treatment arm involved explicit performance-linked incentives. Teachers in the teacher-training arm (T1) did not receive bonuses, promotions, or financial rewards tied to participation or implementation, and external instructors in the external-instructor arm (T2) were employed on fixed contracts that did not vary with student outcomes or measured performance. As a result, differences across treatment arms do not reflect variation in formal incentive schemes. Instead, any differences in implementation intensity are more plausibly attributable to differences in delivery roles, contractual arrangements, and institutional context. Government teachers operate within a civil-service setting characterized by substantial job security and competing instructional and administrative responsibilities, whereas external instructors are hired specifically to deliver the curriculum and can focus primarily on implementation.

3.6 Implementation Fidelity and Monitoring

Monitoring and fidelity checks were embedded throughout the intervention to ensure accurate, comparable, and verifiable implementation across both treatment arms. All T1 and T2 schools were subjected to unannounced visits by external evaluators to document implementation and classroom practices. The evaluators used standardized observation and verification protocols developed jointly by the research and implementation teams (see Appendix A.6 for details).

In T1 schools, fidelity monitoring reflected the flexible and teacher-driven nature of classroom implementation. Because the timing of classroom activities was not fixed or pre-announced, evaluators relied on multiple complementary monitoring approaches. During the initial weeks following the first training round, the field team conducted spot checks to verify whether laboratory materials had been received and were being used. In addition, T1 teachers shared photographic and video documentation of classroom activities through the WhatsApp group. Each submission was required to include the school name, activity title, and date in the accompanying message. This media was compiled into a qualitative dataset to document whether and how specific experiments were implemented. Finally, two rounds of unannounced midline classroom observations conducted during regular science classes served as random checks for T1 schools (see Appendix A.6.3).

In T2 schools, fidelity monitoring relied on direct classroom observations conducted during facilitator-led sessions. Based on facilitator schedules shared by Ecoprism, the DAI data collection team conducted random spot checks using an adapted version of the Stallings classroom observation method, with enumerators recording observations at five-minute intervals using a structured questionnaire. These observations verified facilitator attendance and whether scheduled sessions were conducted, and also captured information on facilitator performance, student engagement with experiments, and use of laboratory materials. In addition, student attendance at facilitator-led sessions was recorded by the facilitator at the beginning or end of each session, and facilitators completed a self-reported digital monitoring tool capturing session timing, student interactions, and other implementation details, which was shared with the field team to corroborate findings from spot checks.

Finally, monitoring data were reviewed through fortnightly calls conducted at the block level with participation from teachers, headmasters, and block-level officers. During these calls, teachers reported progress using tracking sheets documenting activities implemented during the preceding period. These calls facilitated monitoring, coordination, and identification of implementation bottlenecks. Taken together, these systems provided a structured approach to documenting implementation fidelity and capturing process variation across the two treatment models (see Appendix A.6).

4 Descriptive Statistics, Baseline Balance and Statistical Power

4.1 Descriptive Statistics

The intervention focuses on middle school students in grades 6 to 8 who are enrolled in government schools in the four blocks of Fatehpur. These schools face severe constraints on both infrastructure and instructional resources, including limited access to information and communication technology (ICT). The baseline survey indicates that 52% of enrolled students are female, with a similar share (53%)

observed in classrooms during the survey period (see Table 1). The average class size is 30 students, with an average attendance of about 16 per class. Only 41% of schools report having a library facility. Although the average school has nearly three toilets designated for girls, the sanitation infrastructure remains uneven and often inadequate. Taken together, these indicators depict an under-resourced and gender-imbalanced learning environment, consistent with conditions observed elsewhere in rural UP, where basic schooling facilities remain fragile and public investments in learning environments have not kept pace with enrollment expansion.

Table 1: Summary Statistics — School Characteristics

	Mean	SD	Min	Max
	(1)	(2)	(3)	(4)
Infrastructure Index	0	1	-2.534	2.294
ICT Index	0	1	-1.199	2.596
Number of Toilets (Boys)	3.913	1.537	0	9
Number of Toilets (Girls)	3.613	1.545	0	8
Has Library	0.407	0.493	0	1
Mean Enrollment	30.349	12.758	7.667	77.667
Share of Girls Enrolled	0.518	0.066	0.354	0.753
Mean Attendance	16.223	6.878	4.667	41.333
Share of Girls Attending	0.530	0.084	0.273	0.761

Notes: All variables are measured at the school level (N=150). Data on all variables except attendance are obtained from the UDISE database. Enrollment is computed based on the number of students present during the baseline survey. The sample includes all schools with non-missing information in the baseline dataset. For each variable, the table reports the mean, standard deviation (SD), minimum, and maximum values. The infrastructure and ICT (information and communications technology) indices are constructed following the methodology of Anderson (2008). The infrastructure index is based on school-level characteristics from UDISE, including the total number of building blocks and classrooms, the proportion of pucca (permanent) structures, and the availability of electricity. The ICT index is derived from the number of desktop computers and indicators for the presence of a computer lab, internet connectivity, and audiovisual facilities.

The baseline student sample comprises approximately 7,200 participants distributed almost evenly across grades 6, 7, and 8. Among them, 47% are male and the average age is about 12 years. Roughly 29% of students attend private tuition outside of school hours (see Table 2). Students typically belong to relatively large households, with an average size exceeding six members, and often have at least one sibling who assists with homework. The occupational profile of parents highlights the socioeconomic constraints faced by these households. Among mothers, 67% are engaged in domestic work and 10% are engaged in agriculture, while among fathers, 15% hold salaried positions, 12% are self-employed, and 55% work as farmers or laborers. These figures indicate that the majority of families fall into the lower to lower-middle income brackets, depending heavily on informal or seasonal employment. In addition to economic vulnerability, clear gendered patterns emerge in labor participation: women are overwhelmingly concentrated in unpaid household and agricultural work, while men are more diversified across self-employment and wage labor.

Table 2: Summary Statistics — Child Characteristics

	N	Mean	SD	Min	Max
	(1)	(2)	(3)	(4)	(5)
Student in Class 6	7185	0.335	0.472	0	1
Student in Class 7	7185	0.339	0.473	0	1
Student in Class 8	7185	0.326	0.469	0	1
Student is Male	7185	0.468	0.499	0	1
Student Age	7185	11.984	1.268	7	19
Receives Private Tuition	7185	0.288	0.453	0	1
PASEC Literacy Score	7185	1.042	0.776	0	2
Social Desirability Index	7185	0	1	-3.322	3.264
Household size	6819	6.139	1.669	2	11
Siblings helping with homework	6778	1.525	1.130	0	5
Father Works Salaried Job	7185	0.150	0.357	0	1
Father Self-Employed	7185	0.125	0.330	0	1
Father Daily Laborer	7185	0.332	0.471	0	1
Father Works in Farming	7185	0.222	0.415	0	1
Mother Works in Domestic Service	7185	0.674	0.469	0	1
Mother Works in Farming	7185	0.101	0.302	0	1
At Least One Parent Passed Away	7185	0.055	0.228	0	1
Father Completed Primary	7185	0.247	0.431	0	1
Father Completed Secondary	7185	0.299	0.458	0	1
Father Completed College	7185	0.028	0.165	0	1
Mother Completed Primary	7185	0.274	0.446	0	1
Mother Completed Secondary	7185	0.184	0.388	0	1

Notes: All variables are measured at the student (child) level. Data are drawn from self-reported responses collected during the student baseline survey. For each variable, the table reports the number of observations, mean, standard deviation (SD), minimum, and maximum values. The number of observations for household size and the number of siblings who help with homework differs from other variables because a few students reported implausibly large values; these responses were set to missing. The social desirability index is constructed following the methodology of Anderson (2008). Items used to construct this index are listed in Appendix Table A15. The literacy score is based on two literacy questions from the CONFEMEN Programme for the Analysis of Education Systems (PASEC) 2014, adapted into Hindi by Patel and Sandefur (2020), who kindly shared the translations. Each correct response was awarded one point, for a maximum possible score of two. The items for this assessment are presented in Appendix Table A4.

Parental education mirrors these socioeconomic patterns and reveals pronounced gender disparities in educational attainment. More than half (55%) of fathers have completed primary or secondary schooling, though only about 3% hold a college degree. Among mothers, 27% have completed primary and 18% secondary education, with much fewer progressing to higher levels. The gender gap is striking: 30% of fathers versus 18% of mothers have completed secondary education, and overall, about 97% of parents have not pursued college or higher studies. This educational disadvantage underscores the limited availability of role models for learning at home and the intergenerational barriers shaping students' academic trajectories (see Table 2). Together, these demographic and socioeconomic indicators reflect a setting in which children's opportunities to learn—both within and outside the classroom—are constrained by poverty, limited parental education, and structural inequities in the

Table 3: Summary Statistics — Child Outcomes

	Mean	SD	Min	Max
	(1)	(2)	(3)	(4)
Curiosity Index	0	1	-3.057	1.282
Critical Thinking Index	0	1	-3.364	1.439
Growth Mindset Index	0	1	-4.003	2.852
Active Classroom Index	0	1	-3.931	3.961
Aspiration Index	0	1	-3.103	3.220
Raven's Test Score	2.565	1.459	0	5
ASER Math Score	1.571	1.189	0	3
Science Test Score	1.473	1.033	0	4

Notes: All variables are measured at the student (child) level, with 7,185 distinct observations across 150 schools. The sample includes all students with available information in the baseline dataset. Data are drawn from self-reported responses collected during the student baseline survey. For each variable, the table reports the mean, standard deviation (SD), minimum, and maximum values. All indices are constructed following the methodology of Anderson (2008). Items used to construct each index are listed in the following Appendix tables: Table A1 (curiosity, critical thinking, and growth mindset indices) and Table A5 (aspiration index). The Raven's test score is based on five colored progressive matrices from Raven (1965). Each correct response was awarded one point, for a maximum possible score of five. These items were selected from a larger pool using Item Response Theory on pilot data. The corresponding items are listed in Appendix Table A3. The ASER math score is derived from two subtraction items and one adapted multiplication item from ASER Centre (2018). Each correct response was awarded one point, for a maximum possible score of three. The multiplication item was modified to ensure that the correct answer does not include a remainder. The relevant items are presented in Appendix Table A4. The science test score is based on four questions adapted from annual school examinations in the districts of Fatehpur and Sant Kabir Nagar. Students received grade-appropriate items, with one point awarded per correct response (maximum score of four). The full set of questions appears in Appendix Table A2.

school system.

The baseline survey also included several outcome measures designed to capture students' non-cognitive attributes and learning performance. These include indices of curiosity, growth mindset, critical thinking disposition, and educational aspirations, as well as an *active classroom index* reflecting students' perceptions of supportive teaching practices and engagement during science classes. All indices are normalized to have a mean of zero and a range between 0 and 1. Table 3 summarizes these measures alongside students' performance on standardized assessments, including Raven's Progressive Matrices (RPM), ASER tests in mathematics, literacy questions adapted from PASEC (Patel and Sandefur, 2020), and science questions drawn from local annual exams and textbooks. On average, students answered more than half of the RPM items correctly (mean = 2.565 out of 5). For ASER math, the mean score was 1.571 out of 3, while the average literacy score was 2.939 out of 4. The science test, based on a four-point scale, yielded a lower mean score of 1.473, indicating relatively weaker performance in science compared to mathematics and literacy.[7] These results align with broader evidence from UP, which points to persistent gaps in conceptual understanding and scientific reasoning among middle school students despite near-universal enrollment.

[7] All assessment questions except science were common across grades; science questions were grade-specific to ensure age-appropriate content.

Table 4: Summary Statistics — Teacher Characteristics

	Mean	SD	Min	Max
	(1)	(2)	(3)	(4)
Teacher Age	46.53	7.76	28	61
Teacher is Male	0.51	0.50	0	1
Years Teaching in Public Schools	9.73	7.87	0	33
Head Teacher	0.29	0.45	0	1
Science is Main Subject	0.68	0.47	0	1
Ever Attended Training	0.59	0.49	0	1
Teacher Curiosity Index	0	1	-4.48	1.19
Teaching Philosophy Index	0	1	-3.00	4.05
Teacher Locus of Control Index	0	1	-2.22	3.68
Active Learning Practices Index	0	1	-1.14	2.45
Teacher Social Desirability Index	0	1	-3.52	1.04
Teacher Science Test Score	10.24	1	6	12

Notes: Variables are measured at the teacher level, with 160 distinct observations across 150 schools. The sample includes all teachers with available information in the baseline dataset. All data is taken from the teacher baseline survey. The table reports, for each variable, the mean, standard deviation (SD), minimum, and maximum. All indices are constructed using the methodology proposed by Anderson (2008). Items for each index can be found in the following Appendix tables: Table A8 (curiosity index); Table A11 (teaching philosophy, teacher locus of control, and active learning practices); Table A15 (social desirability index). The science test score is based on nine items, adapted from annual exams in the districts of Fatehpur and Sant Kabir Nagar. Three items were taken from each of the grade-specific science questions used in the student baseline survey. Each correct response was assigned one point, for a maximum possible score of nine points. Items for this test can be found in the Appendix: Table A2.

The teacher survey complements these findings by providing information on educators responsible for teaching science in grades 6 through 8. The sample includes 160 teachers with an average age of 46.5 years (ranging from 28 to 61) and a balanced gender composition (51% male). On average, teachers report nearly 10 years of experience in public schools, although the variation is wide, ranging from new recruits to those with more than three decades of service. Approximately 29% hold head teacher positions and 68% identify science as their main subject. Approximately 59% have attended some form of professional development or training, suggesting moderate but uneven access to capacity-building opportunities. The survey also includes indices that capture teacher curiosity, teaching philosophy, locus of control, active learning practices, and social desirability, alongside a science content knowledge test. The average score on this 12-item test was 10.2, indicating relatively strong subject mastery, but leaving open questions about how this knowledge translates into classroom practice. The descriptive statistics for all variables at the teacher-level are presented in Table 4.

4.2 Balance Tables

We test the success of randomization by examining baseline balance between the treatment and control groups using data collected prior to intervention. Because the study design includes two treatment arms, we conduct pairwise comparisons between Treatment 1 and the control group, Treatment 2 and the control group, and between the pooled treatment sample and the control group. This approach allows us to verify that random assignment produced statistically similar groups on observable characteristics and that no systematic differences existed before the intervention. Establishing balance at baseline is essential for ensuring that any post-intervention differences in outcomes can be credibly attributed to the intervention rather than to pre-existing variation across schools, teachers, or students.

Tables 5 and Table 6 report balance tests on school and teacher-level characteristics, respectively, including measures of infrastructure, class size, and teacher experience. Table 7 presents balance in student and household characteristics, such as gender composition, parental education, occupation, and household size, while Table 8 reports baseline values for key outcome variables. Across the 51 variables examined, only four show statistically significant differences at the 10% level. This high degree of balance across the vast majority of characteristics indicates that the randomization was largely successful in generating comparable treatment and control groups. The minimal imbalance observed is consistent with what would be expected by chance given the number of tests conducted. Taken together, these results lend strong support to the internal validity of our design, providing confidence that subsequent treatment effects reflect causal impacts of the intervention rather than baseline disparities.

Table 5: Balance Tables — School Characteristics

	Control (C)	T1	T2	Pooled	*p*-values of difference		
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	T1 - C	T2 - C	Pooled - C
	(1)	(2)	(3)	(4)	(5)=(2)-(1)	(6)=(3)-(1)	(7)=(4)-(1)
Infrastructure Index	0.080	-0.039	-0.041	-0.040	0.535	0.536	0.444
	(0.778)	(1.112)	(1.091)	(1.096)			
ICT Index	-0.163	0.173	-0.011	0.081	0.108	0.440	0.168
	(1.006)	(1.036)	(0.947)	(0.992)			
Number of Toilets (Boys)	3.960	3.860	3.920	3.890	0.727	0.882	0.765
	(1.414)	(1.666)	(1.550)	(1.601)			
Number of Toilets (Girls)	3.640	3.500	3.700	3.600	0.652	0.859	0.881
	(1.675)	(1.432)	(1.542)	(1.484)			
Has Library	0.380	0.420	0.420	0.420	0.681	0.707	0.646
	(0.490)	(0.499)	(0.499)	(0.496)			
Mean Enrollment	29.640	29.820	31.587	30.703	0.956	0.478	0.625
	(10.644)	(12.267)	(15.128)	(13.731)			
Share of Girls Enrolled	0.518	0.521	0.515	0.518	0.778	0.870	0.933
	(0.058)	(0.077)	(0.064)	(0.070)			
Mean Attendance	16.087	16.163	16.420	16.292	0.907	0.791	0.824
	(7.046)	(6.727)	(6.991)	(6.827)			
Share of Girls Attending	0.527	0.530	0.533	0.532	0.838	0.714	0.734
	(0.079)	(0.080)	(0.095)	(0.087)			
N	50	50	50	100			

Notes: Means are reported with standard deviations in parentheses. *p*-values are obtained from ordinary least squares (OLS) regressions with block fixed effects. Standard errors are clustered at the school level (UDISE code). *Pooled* combines Treatment 1 and Treatment 2 and equals 1 if a school received either treatment.

Table 6: Balance Tables — Teacher Characteristics

	Control (C)	T1	T2	Pooled	*p*-values of difference		
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	T1 - C	T2 - C	Pooled - C
	(1)	(2)	(3)	(4)	(5)=(2)-(1)	(6)=(3)-(1)	(7)=(4)-(1)
Teacher Age	46.712	46.111	46.759	46.435	0.700	0.945	0.804
	(8.351)	(6.804)	(8.184)	(7.497)			
Teacher is Male	0.481	0.556	0.481	0.519	0.443	0.919	0.617
	(0.505)	(0.502)	(0.504)	(0.502)			
Years Teaching in Public Schools	9.327	11.333	8.500	9.917	0.265	0.528	0.728
	(8.406)	(8.705)	(6.127)	(7.626)			
Head Teacher	0.308	0.185	0.370	0.278	0.111	0.548	0.603
	(0.466)	(0.392)	(0.487)	(0.450)			
Science is Main Subject	0.673	0.704	0.648	0.676	0.791	0.726	0.961
	(0.474)	(0.461)	(0.482)	(0.470)			
Ever Attended Training	0.558	0.593	0.611	0.602	0.732	0.733	0.697
	(0.502)	(0.496)	(0.492)	(0.492)			
Teacher Curiosity Index	0.048	-0.080	0.034	-0.023	0.555	0.974	0.717
	(1.004)	(1.126)	(0.867)	(1.002)			
Teaching Philosophy Index	-0.015	-0.043	0.057	0.007	0.967	0.698	0.850
	(1.043)	(1.043)	(0.926)	(0.983)			
Teacher Locus of Control Index	0.046	-0.107	0.062	-0.022	0.372	0.870	0.663
	(0.882)	(1.078)	(1.035)	(1.055)			
Active Learning Practices Index	0.078	-0.129	0.054	-0.037	0.343	0.934	0.567
	(1.030)	(0.983)	(0.994)	(0.988)			
Teacher Social Desirability Index	0.045	-0.078	0.035	-0.022	0.560	0.959	0.713
	(0.935)	(1.055)	(1.018)	(1.033)			
Teacher Science Test Score	10.154	10.241	10.315	10.278	0.578	0.482	0.486
	(1.211)	(0.930)	(0.843)	(0.884)			
N	52	54	54	108			

Notes: Means are reported with standard deviations in parentheses. *p*-values are obtained from ordinary least squares (OLS) regressions with block fixed effects. Standard errors are clustered at the school level (UDISE code). *Pooled* combines Treatment 1 and Treatment 2 and equals 1 if a school received either treatment.

Table 7: Balance Tables — Child Characteristics

	Control (C)	T1	T2	Pooled	*p*-values of difference		
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	T1 - C	T2 - C	Pooled - C
	(1)	(2)	(3)	(4)	(5)=(2)-(1)	(6)=(3)-(1)	(7)=(4)-(1)
Student in Class 6	0.325	0.352	0.328	0.340	0.078	0.957	0.297
	(0.468)	(0.478)	(0.470)	(0.474)			
Student in Class 7	0.344	0.338	0.335	0.336	0.725	0.575	0.599
	(0.475)	(0.473)	(0.472)	(0.473)			
Student in Class 8	0.331	0.310	0.337	0.324	0.086	0.623	0.533
	(0.471)	(0.463)	(0.473)	(0.468)			
Student is Male	0.472	0.466	0.466	0.466	0.707	0.724	0.660
	(0.499)	(0.499)	(0.499)	(0.499)			
Student Age	11.964	12.024	11.965	11.994	0.518	0.959	0.724
	(1.269)	(1.280)	(1.256)	(1.268)			
Receives Private Tuition	0.267	0.325	0.273	0.299	0.045	0.847	0.134
	(0.443)	(0.468)	(0.445)	(0.458)			
PASEC Literacy Score	1.060	1.066	1.002	1.034	0.993	0.198	0.455
	(0.768)	(0.779)	(0.779)	(0.780)			
Social Desirability Index	0.034	-0.017	-0.016	-0.017	0.251	0.367	0.252
	(1.012)	(1.005)	(0.982)	(0.994)			
Household size	6.107	6.227	6.084	6.155	0.210	0.513	0.673
	(1.663)	(1.672)	(1.669)	(1.672)			
Siblings helping with homework	1.512	1.509	1.555	1.532	0.850	0.414	0.550
	(1.142)	(1.111)	(1.136)	(1.124)			
Father Works Salaried Job	0.143	0.147	0.160	0.154	0.942	0.292	0.520
	(0.350)	(0.354)	(0.367)	(0.361)			
Father Self-Employed	0.133	0.119	0.122	0.120	0.416	0.456	0.363
	(0.340)	(0.324)	(0.327)	(0.325)			
Father Daily Laborer	0.334	0.334	0.329	0.331	0.813	0.861	0.950
	(0.472)	(0.472)	(0.470)	(0.471)			
Father Works in Farming	0.214	0.233	0.218	0.225	0.367	0.913	0.574
	(0.410)	(0.423)	(0.413)	(0.418)			
Mother Works in Domestic Service	0.676	0.689	0.658	0.673	0.829	0.618	0.875
	(0.468)	(0.463)	(0.474)	(0.469)			
Mother Works in Farming	0.094	0.091	0.118	0.105	0.960	0.147	0.405
	(0.292)	(0.287)	(0.323)	(0.306)			
At Least One Parent Passed Away	0.062	0.050	0.054	0.052	0.101	0.255	0.114
	(0.241)	(0.217)	(0.226)	(0.222)			
Father Completed Primary	0.254	0.252	0.235	0.244	0.894	0.297	0.559
	(0.435)	(0.434)	(0.424)	(0.429)			
Father Completed Secondary	0.281	0.312	0.304	0.308	0.173	0.205	0.118
	(0.450)	(0.463)	(0.460)	(0.462)			
Father Completed College	0.025	0.033	0.026	0.029	0.371	0.718	0.444
	(0.156)	(0.178)	(0.160)	(0.169)			
Mother Completed Primary	0.277	0.263	0.281	0.272	0.386	0.804	0.783
	(0.448)	(0.440)	(0.450)	(0.445)			
Mother Completed Secondary	0.174	0.198	0.180	0.189	0.167	0.634	0.264
	(0.379)	(0.399)	(0.384)	(0.391)			
N	2363	2395	2427	4822			

Notes: Means are reported with standard deviations in parentheses. *p*-values are obtained from ordinary least squares (OLS) regressions with block fixed effects. Standard errors are clustered at the school level (UDISE code). *Pooled* combines Treatment 1 and Treatment 2 and equals 1 if a school received either treatment.

Table 8: Balance Tables — Child Outcomes

	Control (C)	T1	T2	Pooled	*p*-values of difference		
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	T1 - C	T2 - C	Pooled - C
	(1)	(2)	(3)	(4)	(5)=(2)-(1)	(6)=(3)-(1)	(7)=(4)-(1)
Curiosity Index	0.038	-0.008	-0.029	-0.019	0.442	0.297	0.314
	(1.024)	(0.969)	(1.006)	(0.988)			
Critical Thinking Index	0.035	-0.025	-0.010	-0.017	0.292	0.468	0.326
	(1.016)	(0.963)	(1.019)	(0.992)			
Growth Mindset Index	0.030	0.012	-0.041	-0.015	0.642	0.224	0.340
	(1.017)	(1.002)	(0.980)	(0.991)			
Active Classroom Index	-0.023	0.010	0.013	0.011	0.411	0.499	0.382
	(0.979)	(0.995)	(1.025)	(1.010)			
Aspiration Index	-0.015	0.007	0.007	0.007	0.808	0.615	0.671
	(1.008)	(1.000)	(0.993)	(0.996)			
Raven's Test Score	2.621	2.561	2.515	2.538	0.455	0.311	0.319
	(1.483)	(1.446)	(1.447)	(1.447)			
ASER Math Score	1.623	1.605	1.487	1.546	0.851	0.178	0.385
	(1.178)	(1.194)	(1.191)	(1.194)			
Science Test Score	1.490	1.489	1.442	1.465	0.996	0.504	0.672
	(1.051)	(1.026)	(1.020)	(1.023)			
N	2363	2395	2427	4822			

Notes: Means are reported with standard deviations in parentheses. *p*-values are obtained from ordinary least squares (OLS) regressions with block fixed effects. Standard errors are clustered at the school level (UDISE code). *Pooled* combines Treatment 1 and Treatment 2 and equals 1 if a school received either treatment.

4.3 Statistical Power

Table 9 reports the minimum detectable effect size (MDES) calculations for all primary student outcomes, comparing each intervention arm with the control group and examining pooled treatment effects. We present MDES values for our core outcomes—*Curiosity Index*, *Critical Thinking Index*, *Growth Mindset Index*, *Active Classroom Index*, *Aspiration Index*, *Raven's Test Score*, *ASER Math Score*, *Literacy Score*, and *Science Test Score*—across three key contrasts: (i) teacher training versus control, (ii) external instructor versus control, and (iii) both treatment arms pooled versus control.

Table 9: Minimum Detectable Effect Size

	MDE T1 – C	MDE T2 – C	MDE Pooled – C	Control Mean	ICC	N(C)	N(T1)	N(T2)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Curiosity Index	0.186	0.186	0.146	0.038	0.075	2363	2395	2427
Critical Thinking Index	0.180	0.180	0.142	0.035	0.070	2363	2395	2427
Growth Mindset Index	0.154	0.154	0.121	0.030	0.046	2363	2395	2427
Active Classroom Index	0.161	0.161	0.127	-0.023	0.052	2363	2395	2427
Aspiration Index	0.139	0.139	0.110	-0.015	0.034	2363	2395	2427
Raven's Test Score	0.180	0.180	0.141	2.621	0.069	2363	2395	2427
ASER Math Score	0.227	0.227	0.179	1.623	0.121	2363	2395	2427
Literacy Score	0.162	0.162	0.128	2.942	0.053	2363	2395	2427
Science Test Score	0.189	0.189	0.148	1.490	0.078	2363	2395	2427

Notes: The treatment and control arms each include 50 clusters (150 total). All standard errors are clustered at the school level. Indices are constructed following the methodology of Anderson (2008) and standardized using the control group mean and standard deviation (detailed in Appendix A.2). *N* denotes the number of observations. Column (1) reports the minimum detectable effect size (MDES) for the comparison between Treatment 1 (teacher training arm) and the Control Arm. Column (2) reports the MDES for the comparison between Treatment 2 (external instructor arm) and the Control Arm. Column (3) reports the MDES for the comparison pooling both treatment arms relative to the Control Arm. Column (4) reports the mean of the control group. MDES in Columns (1)–(3) are normalized by the standard deviation of the corresponding outcome in the control group. Column (5) reports the intracluster correlation coefficient (ICC). Columns (6)–(8) report the number of observations in the control group, Treatment 1, and Treatment 2, respectively.

To provide complete transparency and context for our power calculations, Table 9 also reports the mean of the control group, the intra-group correlation coefficient (ICC), and the corresponding sample size for each outcome. For instance, we are powered to detect an effect size of approximately 0.186 standard deviations when comparing either treatment arm individually to the control group, and 0.146 standard deviations when pooling both treatment arms, for the *Curiosity Index*. The sample sizes are large and balanced between groups—typically exceeding 2,300 students per arm—thereby enhancing the statistical power and reliability of our design.

All MDES values are expressed in standard deviation units, normalized by the within-control-group variation of each outcome. These calculations account for clustering at the school level, reflecting the unit of randomization, and all indices are constructed following Anderson (2008), using standardized z-scores based on control group parameters. Collectively, these estimates indicate that our study is sufficiently powered to detect moderate and educationally meaningful effects across both academic and psychosocial dimensions, thus strengthening the interpretive credibility of the treatment effects reported in subsequent sections.

5 Data Collection, Outcomes, and Hypotheses

5.1 Implementation Protocol

Baseline data collection was conducted in government schools with the assistance of trained enumerators and field supervisors. Prior permission was obtained from school authorities before scheduling each visit, and the availability of students, teachers, and principals was confirmed in advance. All necessary administrative approvals were secured from the District Magistrate, block-level education officers, and headmasters. Informed consent was obtained from parents several days prior to the scheduled survey. Further details on survey logistics, team composition, and field operations are provided in Section A.3.

5.2 Baseline Survey

The baseline survey was administered over a three-week period during July and August 2025, covering 7,185 students across 150 schools. The surveys were self-administered by students on digital tablets under the supervision of trained enumerators, who provided guidance and clarification as needed. The questionnaire comprised modules on demographics, parental background, curiosity, aspirations, growth mindset, critical thinking, classroom participation, and grade-specific science knowledge. Details of measurement instruments and index construction are provided in Section A.1.

5.3 Administrative Data

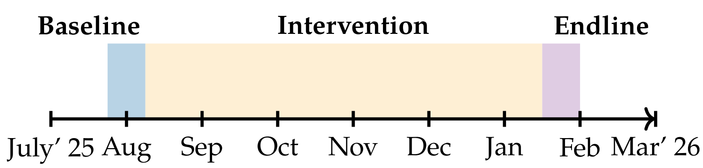
In addition to the survey data, we obtained approval from the District Magistrate to access administrative records from all schools in the sample. These records include student enrollment figures, teacher rosters, teacher attendance, and school infrastructure indicators maintained by local education offices. The administrative data will be used to assess baseline balance and sample representativeness and to examine teacher attendance, entry, exit, and transfers during the intervention period. These measures complement the survey-based outcomes on teacher behavior described in Section 5.6.3.

5.4 Data on Trainers

Because the delivery of the intervention in Treatment 2 relies on external instructors, individual trainer heterogeneity could potentially influence program quality and outcomes. To account for this, we collect detailed trainer-level information on demographics, educational qualifications, and previous work experience—particularly in science instruction—as well as exposure to pedagogical training. Equivalent information is collected for the master trainers responsible for conducting teacher-training workshops in Treatment 1. These data, described in Section A.5, enable us to control for instructor-level variation and to assess the consistency of implementation quality across schools.

5.5 Endline Survey

The endline survey is scheduled for January 2026, following the completion of all planned intervention sessions. This timing ensures that both learning outcomes and attitudinal measures capture students' full exposure to the program while allowing field operations to conclude before the end of the academic year in February 2026.



5.6 Outcomes

This section outlines the primary and secondary outcomes that form the basis of our analysis. We distinguish between direct outcomes—capturing immediate changes in curiosity, engagement, and cognitive skills—and downstream outcomes that reflect broader shifts in learning and psychosocial development. Detailed definitions of each outcome variable, along with the corresponding measurement procedures, are provided in the Appendix A.1. For most measures, we compute simple averages across relevant scale components, whereas for multi-dimensional constructs, we aggregate standardized components following the methodology proposed by Anderson (2008). A comprehensive description of the index construction process appears in Appendix A.2.

The survey instruments draw upon a set of validated scales from psychology and education research. These scales originally employed both five- and seven-point Likert-type response formats. Based on insights from pilot tests, we standardized response categories across instruments to enhance comprehension and ensure consistency during field administration. Unless otherwise noted, all attitudinal items use a five-point Likert scale ranging from "Strongly Disagree" (1) to "Strongly Agree" (5).[8] Minor wording adjustments were made to ensure age appropriateness and contextual relevance, and redundant items were removed where necessary to reduce respondent fatigue and survey length.[9]

All survey instruments were pre-tested in comparable schools within the district before baseline rollout. This pilot process informed the selection, translation and standardization of the response of the elements, ensuring that all constructs were psychometrically valid and contextually meaningful for middle-school students in rural Uttar Pradesh.

5.6.1 Primary Outcomes

Curiosity: To measure student curiosity, we adapt scale measures from Alan and Mumcu (2024)[10], grounded in the information-deprivation framework of Kashdan et al. (2009) and Loewenstein (1994). This scale was originally designed for primary school children in Turkey, a middle-income setting with similar classroom dynamics. We made minor language adjustments following field piloting to ensure clarity and cultural relevance.

Critical Thinking Disposition: Following Alan and Mumcu (2024), we adapt the Critical Thinking Disposition Scale (CTDS) developed by Sosu (2013), which was validated among university students in the United Kingdom. The CTDS does not directly measure cognitive reasoning or analytical ability; instead, it assesses individuals' openness to information that challenges their prior beliefs, their willingness to reflect on alternative viewpoints, and their tendency toward intellectual humility and skepticism.

Cognitive Skills and Academic Performance: Our ultimate interest lies in whether the intervention improves academic outcomes. Using administrative data, we will examine treatment effects on students' final grades. In addition, we administer a multiple-choice assessment designed to measure both curriculum-specific knowledge and broader cognitive skills. As in many developing-country

[8] 1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, 5 = Strongly Agree.

[9] Some scales originally included redundant items to test internal consistency or capture multiple facets of the same construct; these were omitted only where their exclusion did not compromise the reliability of the measure.

[10] We thank the authors for kindly sharing their survey instruments. Their scale draws from Kashdan et al. (2009).

contexts, exams in Uttar Pradesh primarily assess rote recall. This creates a potential tradeoff: if the intervention strengthens conceptual engagement and inquiry, improvements in higher-order skills may not translate one-to-one into gains on recall-heavy tests in the short run. Moreover, time devoted to curiosity-based activities could, in principle, reduce time spent on conventional exam preparation. Because these tensions are part of the policy-relevant choice of adopting pedagogical reforms within constrained instructional time, we interpret impacts on exam-aligned outcomes alongside our measures of curiosity, critical thinking, and classroom engagement.

At the same time, performance on conventional academic assessments remains central to how learning gains are evaluated by schools and education authorities. Accordingly, the first component of our assessment draws on past annual examinations from Fatehpur and Sant Kabir Nagar districts, as well as textbook-based questions, to measure curriculum-aligned knowledge retention. To complement these measures, our baseline survey also includes an assessment of foundational skills and abstract reasoning, drawing on items from the ASER, Hindi-translated PASEC literacy questions shared by Patel and Sandefur (2020), and Raven's Progressive Matrices (Raven, 1938, 1965, 1983).

5.6.2 Secondary Outcomes

Study Preferences and Aspirations: This module captures students' preferences across school subjects and whether they study collaboratively with classmates. Changes in these responses allow us to assess two dimensions: first, whether the intervention increases students' interest in studying science; and second, whether such changes crowd out interest in other subjects. We also ask students whether they plan to pursue science in secondary school—a forward-looking indicator of aspirations toward STEM-related pathways. Finally, questions on peer study habits allow us to examine whether the intervention promotes collaboration and whether such effects extend beyond science to other subjects.

Growth Mindset: Curiosity often leads learners to encounter uncertainty and setbacks, requiring both adaptability and a belief in one's ability to improve through effort. To capture this psychological dimension, we use the "malleability of beliefs" scale of Alan et al. (2019), adapted from the original Growth Mindset scales developed by (Duckworth et al., 2007; Dweck, 2008). The Alan et al. (2019) version, designed for Turkish elementary students, includes items that contextualize beliefs about intelligence within everyday learning experiences, making it particularly suitable for our target age group.

Student-Reported Classroom Activity: We use students' self-reported survey responses to measure changes in classroom participation, teacher behavior, and instructional practices. We adapt items from multiple large-scale surveys, including the Young Lives Survey (YLS) Secondary School Surveys, the Teaching and Learning International Survey (TALIS), the PISA Teacher Questionnaire, and the Trends in International Mathematics and Science Study (TIMSS). These items capture a range of classroom behaviors such as students' frequency of asking questions or raising hands, teachers' instructional effort and feedback, and the balance between interactive and didactic teaching methods. Agreement with most items indicates a more active and collaborative classroom environment, whereas two items are designed to capture traditional classroom modes—memorization through chanting and individual seat work.

Self-Reported Time Use: Changes in science learning outcomes may reflect adjustments in students' at-home learning effort. To examine this, we collect retrospective time-use data on private tutoring and home study outside school, along with information on household chores and leisure activities to capture competing uses of time. To assess potential spillovers across subjects, we also record subject-specific study time, including time spent studying science relative to other courses.

5.6.3 Teacher-Reported Outcomes

We also administer a separate survey to upper-primary science teachers in all study schools. Although many measures rely on self-reports, several can be cross-validated using the corresponding student responses to parallel items.

Curiosity: Teacher curiosity is measured using a scale parallel to that of the student survey, adapted from Alan and Mumcu (2024). This symmetry enables direct comparison of curiosity-related dispositions across teachers and students, shedding light on the potential transmission of curiosity within classrooms.

Cognitive Skills: Teachers complete a cognitive assessment that mirrors the structure of the student test. Because most science teachers in our context teach all three grades within the same school, they are familiar with the full set of lesson plans and instructional materials used in the intervention. Accordingly, the teacher assessment draws questions from all grade-specific modules to capture mastery of the relevant curriculum and underlying scientific concepts.

Teaching Philosophy: This module comprises three components designed to capture teachers' pedagogical beliefs and practices. The first parallels the student survey's teaching-method items, asking teachers to rate both the perceived effectiveness and frequency of use of various approaches to teaching science. The second component adapts items from the TALIS and PISA teacher questionnaires to elicit broader pedagogical orientations, including agreement with teacher-centered approaches that reflect prevailing instructional norms. The third component combines items from the TALIS and YLS teacher surveys with additional questions inspired by the Malleability and Openness to Change scales from Laajaj and Macours (2019). These items assess teachers' beliefs about their instructional efficacy—the degree to which they view their teaching as influencing student outcomes—and their willingness to adopt innovative pedagogical methods.

Self-Reported Time Use: This module captures how teachers allocate their time within and beyond the classroom. Teachers report the frequency with which they engage in core professional activities, including lesson preparation, student interaction outside formal instruction, and collaboration with colleagues. A complementary set of questions records the proportion of each classroom hour devoted to instructional versus administrative tasks. Items are adapted from the teacher survey instruments cited above to ensure comparability across studies.

Knowledge of Students: Following Ashraf et al. (2023), we include an objective measure of teachers' familiarity with their students as a behavioral proxy for engagement and effort. Teachers are asked to estimate the number of students in their class exhibiting specific attributes that are independently collected from student surveys. We then compare teacher estimates to actual student reports, with the premise that more attentive and engaged teachers will provide more accurate assessments of their students.

5.6.4 Unannounced Classroom Observations

Unannounced or "spot-check" visits will be implemented to capture unbiased, externally observed measures of regular teachers' classroom practices and instructional quality. These visits will be conducted by DAI Research & Advisory Services as part of a blinded external verification process designed to collect unbiased, externally observed data on classroom and school-level conditions across 150 sampled schools. Each school will be visited twice on randomly assigned dates, resulting in a total of 300 visits. Enumerators will arrive approximately thirty minutes before the official school start time and follow standardized, non-intrusive procedures adapted from Ganimian et al. (2025). Each evaluator will receive a centrally generated daily field plan with up to three assigned schools per day and will not be informed of the treatment status of any school, thereby ensuring complete blinding to treatment assignment and minimizing potential bias. Upon arrival, the evaluators will introduce themselves as conducting an independent classroom observation as part of a general study, without revealing project or treatment details. They will quietly enter an ongoing class, sit unobtrusively at the back, and document key indicators of school functioning—including teacher presence and behavior, classroom organization, and student engagement—using a standardized Spot-Check Observation Form. The evaluators will not interact with teachers or students and will not provide feedback to school staff during or after the visit.

Evaluators will record whether teachers are present before, at, or after the scheduled start time, verify reported absences through headmaster interviews, and inspect classrooms for physical evidence of activity such as student charts, experimental materials, or notebooks. Field staff will be independently recruited and trained by DAI Research & Advisory Services, separate from the main data collection team, and will participate in structured orientation sessions covering observation protocols, ethical guidelines, and data submission procedures. Quality assurance will be maintained through centralized randomization of school assignments, GPS and timestamp verification for each completed visit, and periodic supervisory back-checks to confirm accuracy. These unannounced visits are designed to minimize observer effects, ensure data reliability, and generate real-time, externally validated measures of pedagogical practices carried out by regular school teachers across both treatment arms and the control group. In particular, the observations capture (i) implementation of the curiosity pedagogy by treated teachers in T1 schools, and (ii) potential behavioral adjustments of regular teachers in T2 schools where external instructors deliver lessons.

6 Conceptual Framework

We conceptualize our intervention within the framework of education production functions that link school, teacher, and student inputs to the formation of cognitive and non-cognitive skills (Hanushek, 2020; Glewwe et al., 2020). Let A denote students cognitive and non-cognitive skills. The education production function can be expressed as:

$$A = f(S, T, e(\theta), A_0), \quad (1)$$

where S represents school inputs, T denotes teacher quality (including pedagogical practices, effort, and content knowledge), and e captures student effort, which depends on beliefs about returns to effort θ as in Jensen (2010). In this framework, student effort enters the production function as a behavioral input. We prioritize at-home learning effort—measured through self-reported study time, subject-specific study time, and private tuition participation—as our preferred proxy because it captures students' autonomous time allocation outside the classroom, consistent with prior work using time use as a measure of effort (Fredricks et al., 2004; Non and Tempelaar, 2016; Stinebrickner and Stinebrickner, 2008). Measures of in-class engagement are analyzed as complementary outcomes that help contextualize classroom behavior, but are not interpreted as standalone measures of student effort. Finally, A_0 captures prior achievement, which may subsume parental inputs and student ability. The function $f(\cdot)$ is increasing and weakly concave in its arguments.

Within this framework, our experiment tests how alternative delivery mechanisms—teacher-led versus externally delivered instruction—alter the productivity of pedagogical inputs in generating learning and curiosity. The study compares two delivery models of a curiosity-based science pedagogy developed by Agastya International Foundation and adapted for government upper-primary schools of Uttar Pradesh: (i) *T1* (*teacher-training*), in which in-service science teachers receive multi-day, in-person training from our implementation partner, Ecoprism, on how to integrate the curriculum into their classroom instruction, and (ii) *T2* (*external-instructor*), in which specialized Ecoprism facilitators directly deliver ten structured curriculum sessions to students. Both interventions seek to raise student learning by stimulating scientific curiosity and engagement, but they differ fundamentally in the locus of delivery and the channels through which teacher effort, pedagogical skill, and student motivation are affected. In essence, they differ in where the productivity gains originate, either through upgrading the skills and motivation of existing teachers or through substituting specialized external instructors for in-school delivery.

Curiosity as an Endogenous Input. In human capital theory, education generates economic returns primarily through its effect on skills rather than credential accumulation (Hanushek, 2020). We model curiosity—defined as the drive to seek information in response to knowledge gaps (Kashdan et al., 2009; Loewenstein, 1994; Keller et al., 2019; Alan and Mumcu, 2024)—as an endogenous behavioral input into the education production function that enhances students' effective learning effort and persistence. Formally, curiosity C enters the production function by increasing both the efficiency with which effort translates into achievement and the level of effort itself:

$$A = f(S, T, e(C, \theta), A_0, C), \tag{2}$$

where $\frac{\partial f}{\partial C} > 0$, $\frac{\partial e}{\partial C} > 0$, and $\frac{\partial^2 f}{\partial e \partial C} > 0$. To reflect that curiosity can raise the productivity of pedagogical inputs, we also allow $\frac{\partial^2 f}{\partial T \partial C} \geq 0$ (and potentially $\frac{\partial^2 f}{\partial S \partial C} \geq 0$). The proposed pedagogical intervention aims to shift the underlying education production function by increasing C and the efficiency with which teaching inputs translate into learning, inducing deeper cognitive engagement, and transforming the classroom environment into one that rewards inquiry, experimentation, and reflection rather than rote memorization. By stimulating curiosity through active experimentation and guided inquiry, the intervention raises learning output for any given level of inputs.

The curriculum consists of 10 lesson plans developed by Agastya, each of which follows a pedagogical arc. Lessons begin with a "super start" activity, a striking demonstration or provocative question designed to trigger curiosity. With the teacher's guidance, students then perform hands-on experiments using locally available materials (soil samples, turmeric indicators, simple magnets, colored water), followed by collaborative reasoning in small groups and reflective discussion. This pedagogical design is grounded in Self-Determination Theory (Ryan and Deci, 2000; Deci and Ryan, 2000), which posits that autonomy (self-directed exploration), competence (mastery experiences), and relatedness (collaborative engagement) drive motivation and engagement. Group-based experiments and guided discussions create opportunities for autonomous exploration and peer collaboration, thereby activating these psychological levers. In economic terms, the intervention relaxes the constraints on both the *quality* of teaching (through new pedagogical tools) and the *returns to effort* for students, shifting the production frontier outward. Enhanced curiosity is expected to increase learning effort both inside and outside the school, potentially crowding in complementary behaviors such as peer study and more time spent on science-related activities. Together, these mechanisms connect classroom-level engagement with student-level behavioral responses to generate improvements in the productivity of public education investments.

Binding Constraints and Policy Alternatives. Past research has found that teachers in developing countries often have weak pedagogical training and rely heavily on lecture-based, memorization-oriented instruction (Glewwe and Muralidharan, 2016). Recent work has found that alternative teaching approaches targeting student inquiry and a scientific mindset significantly improve student learning compared to the status quo (Nourani et al., 2025; Bando et al., 2019). Our own field observations indicate that despite self-reported data to the contrary, teachers in our sample do not put active learning methods into practice. Moreover, we have found that in many schools, science experimentation kits previously provided by schools are incomplete or even missing.

Even if these limitations can be overcome through training, teachers may face additional bureaucratic and institutional constraints that could reduce a teacher's ability or willingness to implement pedagogical reforms. Although systematic evidence is limited, existing work suggests that public school teachers in developing countries face multiple demands on their time, including administrative and non-teaching tasks (Evans and Yuan, 2018; Kim, 2019). For example, a recent report on public schools in Dehli finds that teachers only spend about half of their time on academic tasks (DCPCR and Accountability Initiative, 2018). In our context, where schools have limited support staff, the administrative burden teachers face is likely to be higher, limiting their ability to prepare and adopt novel teaching methods.

These observations suggest that teachers operate under binding capacity and incentive constraints that prevent them from successfully imparting their knowledge to their students. These constraints can be formalized as a restricted production technology:

$$A^{\text{status quo}} = f(S, T^{\text{low}}, e^{\text{low}}, A_0), \tag{3}$$

where T^{low} denotes low baseline teacher quality and e^{low} reflects weak student engagement, implying that the marginal product of teacher inputs is low. Teachers often lack both the incentives and the know-how to facilitate experimentation or link concepts to daily life, resulting in low engagement and weak conceptual understanding. Because of limited training, materials, and peer support, teachers face high marginal costs of adopting active learning methods and often hold pessimistic beliefs about the returns to pedagogical effort, reinforcing low-quality equilibria. These binding constraints motivate our two distinct policy approaches.

T1 (Teacher-Training). In *T1*, in-service teachers participate in multi-day, in-person training workshops conducted by Ecoprism to learn and practice the new curiosity-based pedagogy. Training enhances their content mastery and introduces active learning methods, emphasizing the psychology of curiosity, inquiry-based learning, and student-centered discussion practices. Teachers are encouraged to apply these techniques beyond the ten pre-planned Agastya lessons and are supplied with all necessary materials to conduct the experiments. The upgraded teacher quality can be expressed as:

$$T^1 = T^{\text{low}} + \delta\tau, \qquad \delta > 0, \tag{4}$$

where τ represents the training input and δ captures training effectiveness. By observing greater student engagement, teachers may update their beliefs about their own efficacy and the returns to effort, shifting their teaching philosophy toward student-centered instruction. This belief-updating mechanism parallels models in which increased teacher knowledge raises the productivity of teaching inputs (Strøm and Falch, 2020). To sustain engagement after the workshop, teachers are added to WhatsApp groups moderated by Ecoprism trainers, where they can seek guidance, troubleshoot, and share classroom experiences. Continuous peer support via these moderated groups, along with the provision of experiment materials, lowers the marginal cost of sustained adoption of the curriculum and supports implementation fidelity.

Critically, this approach invests in the long-run productivity of existing public-sector human capital and the endogenous accumulation of teacher capital, plausibly generating dynamic benefits: teachers who internalize new methods may continue to use them across cohorts, raising the long-run stock of pedagogical capital within schools and generating horizontal spillovers through peer learning. However, implementation fidelity depends on sustained teacher effort, which may be weakened by competing demands, weak monitoring, and organizational constraints within the public school system. The effectiveness of T1 hinges on whether training-induced improvements in teacher quality, pedagogical practices, and beliefs about student potential are large enough and sufficiently durable to meaningfully shift both teacher and student outcomes.

T2 (External-Instructor). In *T2*, specialized Ecoprism facilitators directly deliver the same ten Agastya-designed curriculum sessions to students. Regular classroom instruction continues as usual, and teachers in these schools do not receive additional training. This design circumvents teacher capacity and incentive constraints, enabling greater short-run fidelity and standardization of instruction. During externally delivered sessions, the instructional quality is high:

$$T^{\text{session}} = T^2 \tag{5}$$

Over the entire term, the effective teacher quality under T2 can be written as:

$$\tilde{T}^2 = \frac{s\,T^2 + T^{\text{low}}}{1+s}, \qquad s \geq 0, \tag{6}$$

where s denotes the ratio of external instructional time to regular teacher time (with regular teacher time normalized to one). Any $s > 0$ implies an increase in total instructional time, with larger s corresponding to more intensive augmentation by external sessions.[11]

Post-intervention, regular teachers may learn by observing these facilitators, leading to potential pedagogical spillovers:

$$T^{\text{post}} = T^{\text{low}} + \lambda\, s\,(T^2 - T^{\text{low}}), \qquad \lambda \in [0,1], \tag{7}$$

where λ captures the degree of observational learning or pedagogical spillovers from exposure to external instructors.[12] When λ is small, T2 yields limited long-run gains in teaching quality despite strong short-run effects, whereas higher λ implies more durable improvements as regular teachers internalize elements of the new pedagogy.

Policy Trade-offs and Equilibrium Effects. The two delivery models thus embody a policy-relevant trade-off between *building internal capacity* and *outsourcing specialized delivery* that mirrors implementation choices policymakers typically face when scaling pedagogical reforms in resource-constrained settings (Banerjee et al., 2017; Ganimian, 2020). T1 invests in improving the long-run productivity of existing public-sector inputs and generates potential dynamic spillovers, but faces implementation challenges,

[11] In T2, external instructors *augment* rather than replace regular instruction: regular teachers continue their own classes, while facilitators deliver additional sessions using the same content and pedagogy as in T1. As a result, total instructional time may exceed that in T1. If regular teachers adjust their teaching time or classroom practices after observing facilitators, such behavioral responses could either offset or amplify these differences. We assess these adjustments using unannounced spot-checks and classroom observations that record teacher attendance and instructional practices during regular classes. In this setting, effects on curiosity-related outcomes are most plausibly attributable to the pedagogical quality of facilitator-led sessions, whereas improvements in test scores without corresponding gains in curiosity-related outcomes would be consistent with effects driven primarily by increased instructional time rather than pedagogical innovation (Agüero and Beleche, 2013).

[12] Although such spillovers are theoretically possible, the supplementary lessons are scheduled separately from regular classes and do not require teacher attendance. Field observations suggest that teachers are often engaged in other school duties during these periods, limiting sustained exposure to facilitator-led instruction.

heterogeneous take-up, and weak monitoring. T2 offers better short-term quality assurance and ease of supervision, but entails higher per-student cost, limited scalability, limited persistence, and weaker institutional integration—a margin that is central to debates on service delivery in low-capacity public systems (Banerjee et al., 2017; Glewwe et al., 2020; Muralidharan and Sundararaman, 2013). Policymakers in low-capacity systems frequently confront this trade-off when deciding whether to upgrade teacher quality or to contract external agents to accelerate learning.

Finally, both approaches may generate broader equilibrium effects. T1 could create knowledge spillovers across peers and future cohorts, diffusing improved practices, and altering pedagogical norms across schools. T2 might shift student aspirations, parental perceptions of science learning quality, and household investments in education. In both cases, sustained improvements in curiosity and cognitive engagement can translate into higher accumulation of human capital, increasing the efficiency with which inputs are transformed into skills and potentially influencing future educational choices, occupational outcomes, and labor market productivity.

Our empirical analysis estimates the reduced-form impacts of T1 and T2 on three categories of outcomes. *Primary student outcomes* include curiosity, critical thinking disposition, and cognitive skills measured through curriculum-aligned science tests, ASER mathematics assessments, and Raven's Progressive Matrices. *Secondary student outcomes* capture growth mindset, educational aspirations, student-reported classroom practices, study preferences across subjects, and time allocation to learning activities both inside and outside school. *Teacher outcomes* include teacher curiosity, content knowledge, teaching philosophy, classroom time use, and knowledge of individual students. Additionally, through unannounced ("spot-check") visits across all schools, we will observe regular teachers' attendance and classroom practices to assess the impact of the intervention on teaching behavior.[13] Comparing the relative effectiveness of the intervention across the treatment groups provides policy-relevant evidence on whether to intervene in the education production function by upgrading teacher inputs or contracting temporary high-quality alternatives to improve learning efficiency in low-resource environments.

[13] This will also help us in detecting potential pedagogical spillovers on regular teachers from exposure to external instructors in T2.

7 Empirical Strategy

7.1 Reduced-Form Specification

We estimate intent-to-treat (ITT) effects using the following reduced-form specification:

$$Y_{i,s,t=1} = \beta_0 + \beta_1 T1_s + \beta_2 T2_s + \beta_3 Y_{i,s,t=0} + \beta_4 X_{i,s} + e_{i,s,t}$$

where $Y_{i,s,t=1}$ denotes the endline outcome for individual i from school s, $Y_{i,s,t=0}$ is the corresponding baseline measure, and $X_{i,s}$ is a vector of pre-specified controls. $T1_s$ and $T2_s$ are indicator variables denoting assignment to Treatment 1 (*Teacher-Training Model*) and Treatment 2 (*External-Instructor Model*), respectively. Standard errors are clustered at the school level, the unit of randomization.

7.2 Heterogeneous Impact of the Intervention

We examine heterogeneous treatment effects across several dimensions. First, we assess heterogeneity by student gender, given well-documented gender gaps in learning outcomes in this setting (ASER, 2022; Rakshit and Sahoo, 2023). We also examine heterogeneity by socioeconomic background to understand how the intervention differentially affects students from varied household strata (Das et al., 2022). In addition, we explore heterogeneity related to instructor characteristics across delivery models, recognizing that schoolteachers and external instructors differ along observable dimensions such as age, educational background, teaching experience, and contractual arrangements. While these characteristics are inherently bundled with the delivery model and cannot be causally disentangled, we examine whether treatment effects vary systematically with baseline measures of teacher characteristics, where feasible, to provide descriptive evidence on how instructor attributes may influence observed impacts.

7.3 Measuring Spillover

We consider two potential spillover effects in this study. First, pedagogical training in the teacher-training arm (T1) may generate spillovers across subjects if teachers apply elements of the curiosity-based pedagogy beyond science instruction, or if their peers learn from them. If such learning occurs—through peer interactions or teachers' own instructional adaptation—improvements in teaching practices may translate into gains in student learning in subjects other than science, even though these subjects are not directly targeted by the intervention. To examine this possibility, we use administrative data on student learning outcomes in non-science subjects and compare these outcomes between T1 schools and control schools.

Second, spillovers may arise in the external-instructor arm (T2) if regular teachers observe facilitator-led sessions and subsequently adapt aspects of their own classroom practices. We examine this channel using a combination of midline unannounced spot-check data and self-reported information from teachers and students collected at endline. Specifically, we compare measures of teacher presence, instructional activities, and classroom practices during regular classes in T2 schools and control schools. Systematic differences between these groups would be consistent with spillovers from facilitators to regular teachers operating through observational learning or informal exposure.

Together, these analyses allow us to assess whether pedagogical spillovers operate across subjects in T1 or from facilitators to regular teachers in T2, and to interpret the main treatment effects in light of potential diffusion of pedagogical practices within schools.

7.4 Multiple Hypotheses Testing

For all primary outcome variables, we construct composite indices following the procedure outlined in Appendix A.2. To account for multiple hypothesis testing, we adjust p-values to control the false discovery rate (FDR) within outcome families and report the corresponding q-values (Benjamini and Heller, 2007). These corrections for multiple testing are applied exclusively to the set of primary outcomes.

7.5 Addressing Attrition

We will systematically track attendance by documenting student participation in each intervention session. This continuous monitoring will allow us to incorporate attendance patterns directly into the analysis. Because the intervention is integrated into the regular school schedule and our monitoring protocol involves repeated field visits, we expect the attrition to be minimal. Nevertheless, if we observe differential attrition between the treatment and control groups, we will apply the Lee bounds as a correction method (Kremer et al., 2009; Baird et al., 2011; Drexler et al., 2014; Fiala et al., 2022).

Lee bounds rely on a monotonicity assumption—specifically, that the treatment affects attrition in only one direction—which we believe is satisfied by our study design. However, if endline data indicate a potential violation of this assumption, we will implement alternative correction approaches, including those proposed by Molina Millán and Macours (2017).

7.6 Addressing Social Desirability Bias

To address potential response bias in self-reported measures, we administer the 13-item short form of the Crowne–Marlowe Social Desirability Scale (Crowne and Marlowe, 1960; Reynolds, 1982) to both teachers and students (Table A15). This well-established instrument captures the tendency of respondents to present themselves in a socially favorable light rather than report candidly. Among students, it enables more accurate interpretation of self-reported non-cognitive outcomes such as curiosity, confidence, and growth mindset. Among teachers, it helps adjust for bias in reported motivation, openness to inquiry-based pedagogy, and classroom practices. Because curiosity and related traits are socially valued within the intervention context, this scale helps distinguish genuine internalization of attitudes from impression management. The short form exhibits strong psychometric reliability and has been validated in comparable educational field settings, including recent studies in India (Dhar et al., 2022).

7.7 Addressing Outcomes with Limited Variation

If certain outcome variables exhibit limited variation, we will proceed as follows. First, we will assess whether 95 percent or more of the observations within the treatment group take the same value. If this condition holds, the variable will be excluded from the analysis, including from any composite indices in which it appears. If all constituent variables of an index are excluded on this basis, the index itself will be dropped from the evaluation.

7.8 Addressing other Potential Concerns

A key methodological concern relates to potential cross-contamination, wherein students from different schools may interact and influence each other's academic outcomes or behaviors, including through possible inter-school transfers. However, several factors substantially mitigate this risk in our setting.

First, the intervention is implemented mid–academic year and concludes within the same academic cycle, making school transfers during this period highly unlikely. Second, students typically enroll in geographically proximate schools, which limits the extent of cross-school social networks. Third, we will maintain detailed records of any student entries to or exits from our sample resulting from school changes, enabling us to identify and adjust for any potential bias in our estimates.

8 Administrative Information

8.1 Funding

This project was supported by the J-PAL *Learning for All Initiative*, Northeastern University, and the Building Inclusive Growth (BIG) Lab and the Kellogg Institute for International Studies at the University of Notre Dame. The authors gratefully acknowledge their financial support for this research.

8.2 Institutional Review Board Approval

This study received Institutional Review Board (IRB) approval from DAI Research and Advisory Services and Northeastern University (IRB # 24-07-12).

8.3 Declaration of Interest

The authors declare that they have no known competing financial or other interests that could have influenced the work reported in this paper.

8.4 Acknowledgments

The authors gratefully acknowledge the invaluable cooperation of the District Magistrate of Fatehpur, Mr. Ravinder Singh (Indian Administrative Service), as well as Ananth Nag M., Amitav Nath, school principals, teachers, participating students, and their parents across the study schools, all of whom were exceptionally generous with their time and support. We thank Aaditya Dubey and Vagisha Pandey for outstanding research assistance, and Bhuvnesh Bajpai, Sitaram Mukherjee, Anup Roy, and Manvendra Singh, along with the entire field team at DAI Research and Advisory Services, for their dedication to high-quality data collection. We are also deeply grateful to the Agastya International Foundation for its partnership and guidance in the design of the curriculum.

8.5 Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this manuscript the authors used Claude.AI in order to improve grammar and come up with this creative title. After using Claude.AI, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

Abadzi, H. (2009): "Instructional time loss in developing countries: Concepts, measurement, and implications," *World Bank Research Observer*, 24, 267–290.

Agüero, J. M. and T. Beleche (2013): "Test-Mex: Estimating the effects of school year length on student performance in Mexico," *Journal of Development Economics*, 103, 353–361.

Ahmad, Z., S. Ismail, and N. A. Rahman (2025): "Work Stress in Public Service: Job Demands Outweigh Transformational Leadership," *Asian Journal for Public Opinion Research*, 13, 59–77.

Alan, S., T. Boneva, and S. Ertac (2019): "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit," *The Quarterly Journal of Economics*, 134, 1121–1162.

Alan, S. and S. Ertac (2018): "Fostering patience in the classroom: Results from randomized educational intervention," *Journal of Political Economy*, 126, 1865–1911.

Alan, S. and I. Mumcu (2024): "Nurturing childhood curiosity to enhance learning: Evidence from a randomized pedagogical intervention," *American Economic Review*, 114, 1173–1210.

Anderson, M. L. (2008): "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 103, 1481–1495.

ASER (2022): "Pratham Education Foundation, Annual Status of Education Report (ASER) 2022," India.

ASER Centre (2018): "ASER 2018 Tool (English Version)," Annual Status of Education Report (Rural) learning assessment tool.

——— (2024): "Annual Status of Education Report 2024: Uttar Pradesh Rural Findings," *Annual Status of Education Report*, data based on household survey across 70 districts of Uttar Pradesh.

Ashraf, N., A. Banerjee, and V. Nourani (2023): "Learning to teach by learning to learn," *Ms., University of Chicago and Makerere University*.

Baird, S., C. McIntosh, and B. Özler (2011): "Cash or condition? Evidence from a cash transfer experiment," *The Quarterly journal of economics*, 126, 1709–1753.

Bando, R., P. Gertler, E. Bettinger, M. Kaffenberger, and E. King (2019): "School-Based Management and Learning Outcomes: Experimental Evidence from Brazil," *American Economic Journal: Applied Economics*, 11, 140–168.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shotland, and M. Walton (2017): "From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application," *Journal of Economic Perspectives*, 31, 73–102.

Benjamini, Y. and R. Heller (2007): "False discovery rates for spatial signals," *Journal of the American Statistical Association*, 102, 1272–1281.

Bharti, N. K., S. Malik, A. Mukhopadhyay, and N. Prakash (2024): "Science on the Move: How Experiential Learning Shapes Human Capital," *Working Paper*, working Paper.

Bold, T., M. Kimenyi, G. Mwabu, A. Ng'ang'a, and J. Sandefur (2018): "Experimental evidence on scaling up education reforms in Kenya," *Journal of Public Economics*, 168, 1–20.

Bruns, B. and J. Luque (2014): *Great Teachers: How to Raise Student Learning in Latin America and the Caribbean*, World Bank Publications.

Census (2011): "Data from the Office of the Registrar General & Census Commissioner, India," https://apfstatic.s3.ap-south-1.amazonaws.com/s3fs-public/Champawat.pdf, accessed: 2024-07-29.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and F. H. Rogers (2006): "Missing in action: Teacher and health worker absence in developing countries," *Journal of Economic Perspectives*, 20, 91–116.

Crowne, D. P. and D. Marlowe (1960): "A new scale of social desirability independent of psychopathology," *Journal of Consulting Psychology*, 24, 349–354.

Das, J., A. Singh, and A. Y. Chang (2022): "Test Scores and Educational Opportunities: Panel Evidence from Five Low- and Middle-Income Countries," *Journal of Public Economics*, 208, 104610.

DCPCR and Accountability Initiative (2018): "Report on Time Allocation and Work Perceptions of Teachers: A Look Inside Delhi's Government and Municipal Schools," Delhi Commission for Protection of Child Rights (DCPCR) and Accountability Initiative, Centre for Policy Research, New Delhi.

Deci, E. L. and R. M. Ryan (2000): "The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior," *Psychological Inquiry*, 11, 227–268.

Delfgaauw, J. and R. Dur (2008): "Incentives and workers' motivation in the public sector," *The Economic Journal*, 118, 171–191.

Deming, D. J. (2017): "The growing importance of social skills in the labor market," *The Quarterly Journal of Economics*, 132, 1593–1640.

Dhar, D., T. Jain, and S. Jayachandran (2022): "Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India," *American economic review*, 112, 899–927.

Drexler, A., G. Fischer, and A. Schoar (2014): "Keeping it simple: Financial literacy and rules of thumb," *American Economic Journal: Applied Economics*, 6, 1–31.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly (2007): "Grit: Perseverance and passion for long-term goals," *Journal of Personality and Social Psychology*, 92, 1087–1101.

Duflo, E., P. Dupas, and M. Kremer (2015): "School governance, teacher incentives, and pupil–teacher ratios: Experimental evidence from Kenyan primary schools," *Journal of Public Economics*, 123, 92–110.

Dweck, C. S. (2008): *Mindset: The New Psychology of Success*, New York: Random House Digital, Inc.

Evans, D. K. and F. Yuan (2018): "The Working Conditions of Teachers in Low- and Middle-Income Countries," Policy Research Working Paper 8656, World Bank, Washington, DC.

Fiala, N., A. Garcia-Hernandez, K. Narula, and N. Prakash (2022): "Wheels of change: Transforming girls' lives with bicycles," .

Firestone, W. A. and J. R. Pennell (1993): "Teacher commitment, working conditions, and differential incentive policies," *Review of Educational Research*, 63, 489–525.

Fredricks, J. A., P. C. Blumenfeld, and A. H. Paris (2004): "School engagement: Potential of the concept, state of the evidence," *Review of Educational Research*, 74, 59–109.

Frey, B. S. and R. Jegen (2001): "Motivation crowding theory," *Journal of Economic Surveys*, 15, 589–611.

Fryer, R. G. (2017): "Management and student achievement: Evidence from a randomized field experiment," *American Economic Review*, 107, 2913–2953.

Ganimian, A. J. (2020): "Growth-Mindset Interventions at Scale: Experimental Evidence From Argentina," *Educational Evaluation and Policy Analysis*, 42, 417–438.

Ganimian, A. J., I. M. Mbiti, and A. Mishra (2025): "Leveraging Complementarities between Teachers' Content Knowledge and Pedagogical Supports: Experimental Evidence from India," Working Paper.

Giauque, D., S. Anderfuhren-Biget, and F. Varone (2013): "Stress perception in public organisations: Expanding the job demands–resources model by including public service motivation," *Review of Public Personnel Administration*, 33, 58–83.

Glewwe, P., S. Lambert, and Q. Chen (2020): "Education Production Functions: Updated Evidence from Developing Countries," in *The Economics of Education: A Comprehensive Overview*, ed. by S. Bradley and C. Green, London, UK: Elsevier, 261–304, 2 ed.

Glewwe, P. and K. Muralidharan (2016): "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications," Elsevier, vol. 5 of *Handbook of the Economics of Education*, 653–743.

Hanushek, E. A. (2020): "Education Production Functions," in *The Economics of Education: A Comprehensive Overview*, ed. by S. Bradley and C. Green, London, UK: Elsevier, 161–184, 2 ed.

Haushofer, J. and J. Shapiro (2016): "The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya," *The Quarterly Journal of Economics*, 131, 1973–2042.

Heckman, J. J. and T. Kautz (2012): "Hard evidence on soft skills," *Labour economics*, 19, 451–464.

Jackson, C. K. (2018): "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes," *Journal of Political Economy*, 126, 2072–2107.

Jensen, R. (2010): "The (Perceived) Returns to Education and the Demand for Schooling," *Quarterly Journal of Economics*, 125, 515–548.

Kashdan, T. B., M. W. Gallagher, P. J. Silvia, B. P. Winterstein, W. E. Breen, D. Terhar, and M. F. Steger (2009): "The curiosity and exploration inventory-II: Development, factor structure, and psychometrics," *Journal of Research in Personality*, 43, 987–998.

Keller, U., A. Strobel, R. Wollschläger, S. Greiff, R. Martin, M.-P. Vainikainen, and F. Preckel (2019): "A need for cognition scale for children and adolescents: Structural analysis and measurement invariance," *European Journal of Psychological Assessment*, 35, 137–149.

Kim, K. N. (2019): "Teachers' Administrative Workload Crowding Out Instructional Activities," *Asia Pacific Journal of Education*, 39, 31–49.

Kingdon, G. and M. Muzammil (2009): "A Political Economy of Education in India: The Case of Uttar Pradesh," Tech. rep., University of Cambridge, RECOUP Working Paper No. 22.

Kremer, M., C. Brannen, and R. Glennerster (2013): "The challenge of education and learning in the developing world," *Science*, 340, 297–300.

Kremer, M., E. Miguel, and R. Thornton (2009): "Incentives to learn," *The Review of Economics and statistics*, 91, 437–456.

Kumar, D., N. Sunder, R. Sabates Aysa, and W. Wadhwa (2024): "Improving Children's Foundational Learning through Community–School Participation: Experimental Evidence from Rural India," *Labour Economics*, 91, 102615.

Laajaj, R. and K. Macours (2019): "Measuring Skills in Developing Countries," *Journal of Human Resources*.

Lins de Holanda Coelho, G., P. H. P. Hanel, and L. J. Wolf (2020): "The very efficient assessment of need for cognition: Developing a six-item version," *Assessment*, 27, 1870–1885.

Loewenstein, G. (1994): "The psychology of curiosity: A review and reinterpretation," *Psychological Bulletin*, 116, 75–98.

Loyalka, P., A. Popova, G. Li, and Y. Shi (2019): "Does teacher training actually work? Evidence from a large-scale randomized evaluation of a national teacher training program," *American Economic Journal: Applied Economics*, 11, 128–154.

Ministry of Education (2020): "National Education Policy 2020," .

Molina Millán, T. and K. Macours (2017): "Attrition in randomized control trials: Using tracking information to correct bias," *IZA Discussion Paper*.

Muralidharan, K. (2024): *Accelerating India's Development: A State-Led Roadmap for Effective Governance*, New York: Viking.

Muralidharan, K., A. Singh, and A. J. Ganimian (2019): "Disrupting education? Experimental evidence on technology-aided instruction in India," *American Economic Review*, 109, 1426–1460.

Muralidharan, K. and V. Sundararaman (2013): “Contract Teachers: Experimental Evidence from India,” *Journal of Political Economy*, 121, 273–327.

——— (2015): “The aggregate effect of school choice: Evidence from a two-stage experiment in India,” *The Quarterly Journal of Economics*, 130, 1011–1066.

Non, A. and D. Tempelaar (2016): “Time preferences, study effort, and academic performance,” *Economics of Education Review*, 54, 36–61.

Nourani, V., N. Ashraf, and A. Banerjee (2025): “Learning to Teach by Learning to Learn: Evidence from Uganda,” *Forthcoming*, working paper, University of Chicago and Makerere University.

Patel, D. and J. Sandefur (2020): “A Rosetta Stone for Human Capital,” Working Papers 550, Center for Global Development.

Perkins, D. N. (1992): *Smart Schools: Better Thinking and Learning for Every Child*, New York: Free Press.

Rakshit, A. and S. Sahoo (2023): “Biased Teachers and the Gender Gap in Learning Outcomes,” *Journal of Development Economics*, 163, 103123.

Rasul, I., D. Rogger, and M. J. Williams (2021): “Management, organizational performance, and task clarity: Evidence from Ghana’s civil service,” *Journal of Public Administration Research and Theory*, 31, 259–277.

Raven, J. C. (1938): *Progressive Matrices: A Perceptual Test of Intelligence*, London: H. K. Lewis & Co.

——— (1965): *Guide to Using the Coloured Progressive Matrices, Sets A, Ab, B*, London: H. K. Lewis, (Revised order, 1956).

——— (1983): “Manual for Raven’s progressive matrices and vocabulary scales,” *Standard progressive matrices*.

Reynolds, W. M. (1982): “Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale,” *Journal of Clinical Psychology*, 38, 119–125.

Romero, M., J. Sandefur, and W. A. Sandholtz (2020): “Outsourcing education: Experimental evidence from Liberia,” *American Economic Review*, 110, 364–400.

Ryan, R. M. and E. L. Deci (2000): "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being," *American Psychologist*, 55, 68–78.

Schwab, B., S. Janzen, N. P. Magnan, and W. M. Thompson (2021): "SWINDEX: Stata module to create a standardized weighted index of multiple indicator variables," Statistical Software Components, Boston College Department of Economics, accessed on 2024-08-14.

Skaalvik, E. M. and S. Skaalvik (2017): "Dimensions of teacher burnout: Relations with potential stressors at school," *Social Psychology of Education*, 20, 775–790.

Sorrenti, G. et al. (2024): "University Education and Non-Cognitive Skill Development," *Oxford Economic Papers*, 76, 1–25.

Sosu, E. M. (2013): "The development and psychometric validation of a critical thinking disposition scale," *Thinking Skills and Creativity*, 9, 107–119.

Stinebrickner, R. and T. R. Stinebrickner (2008): "The causal effect of studying on academic performance," *The BE Journal of Economic Analysis & Policy*, 8, 1–55.

Strøm, B. and T. Falch (2020): "The Role of Teacher Quality in Education Production," in *The Economics of Education: A Comprehensive Overview*, ed. by S. Bradley and C. Green, London, UK: Elsevier, 307–319, 2 ed.

Swartz, R. J., D. N. Perkins, et al. (2017): *Thinking-Based Learning: Activating Students' Potential*, Cambridge: Cambridge University Press.

Yoon, K. S., T. Duncan, S. W.-Y. Lee, B. Scarloss, and K. L. Shapley (2007): "Reviewing the evidence on how teacher professional development affects student achievement," *Issues & Answers Report*, 33.

A Appendices

A.1 Outcome Variables

Appendix Table A1: Student Soft Skills Measures

Outcomes	Definitions
Curiosity Index	*Response Options: 1 - Strongly disagree; 5- Strongly agree* 1. Mysterious things which are hard to understand make me curious. 2. It gets on my nerves if I am close to solving a question but can't figure it out fully. 3. I am close to solving a question but can't figure it out fully. I get ambitious. 4. It is frustrating not knowing a question. So, I try working harder to learn it.
Critical Thinking Disposition	*Response Options: 1 - Strongly disagree; 5- Strongly agree* 1. I am often on the lookout for new ideas. For example, by speaking to my friends or family members about new things. 2. Sometimes in class, I hear something from the teacher that makes me think I was wrong about how something works. 3. It's important to understand other people's viewpoints on an issue. 4. It is important to justify the decisions I make to others. 5. I usually think about the wider implications of a decision before taking action. For example, while thinking whether to study or help my younger siblings in studying, I think about the impact of both of these things.
Growth Mindset	*Response Options: 1 - Strongly disagree; 5- Strongly agree* 1. I believe I can expand my intelligence through learning new things. 2. One has a certain amount of intelligence, and one can't really do much to change it. $(-)$ 3. If you're not good at a subject, working hard won't make you good at it. $(-)$ 4. If I study hard enough, I could be the most successful student in the class.

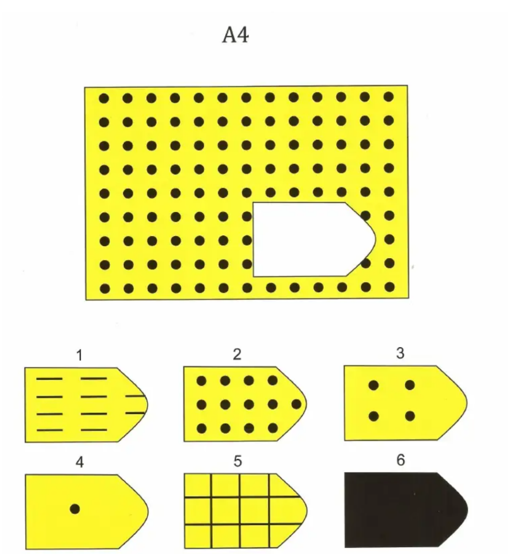
Appendix Table A2: Student Science Questions

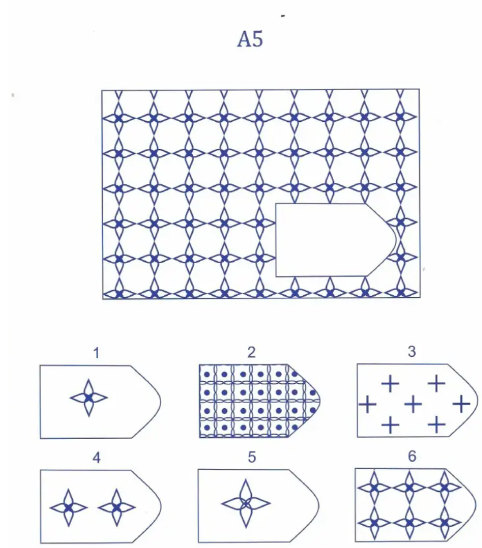
Outcomes	Definitions
Class 6 Science Test	*Multiple-choice responses* 1. How many centimeters are there in a meter? (a) 1000 cm (b) 200 cm (c) 100 cm (d) 10 cm 2. How many parts are there in the human brain? (a) Two (b) Three (c) Four (d) Five 3. The building block of matter is (a) Atom (b) Electron (c) Proton (d) Neutron 4. The motion of a pendulum in a clock is an example of –

	(i) Rectilinear motion (ii) Oscillatory motion (iii) Rotational motion (iv) Circular motion
Class 7 Science Test	*Multiple-choice responses* 1. Which of the following is a chemical change? (a) Bulb glowing (b) Fan running (c) Rusting (d) Ice melting 2. Insectivorous plants are generally found in areas where the soil lacks - (i) Oxygen (ii) Water (iii) Nitrogen (iv) Carbon 3. The boiling point of water is- (a) 10°C (b) 100°C (c) 120°C (d) 40°C 4. Mechanical energy changes to – when two stones strike rapidly (i) Light and sound energy (ii) Heat and sound energy (iii) Heat and light energy (iv) Heat, light, and sound energy
Class 8 Science Test	*Multiple-choice responses* 1. The unit of pressure is - (a) Newton/meter (b) Kilogram (c) Joule (d) Newton/m^2 2. Which of the following is a Rabi crop- (a) Rice (b) Corn (c) Peanuts (d) Wheat 3. Which of the following is attracted to a magnet - (a) Sawdust (b) Glass piece (c) Iron filings (d) Copper filings 4. Which particle has a negative charge? (i) Proton (ii) Electron (iii) Neutron (iv) All

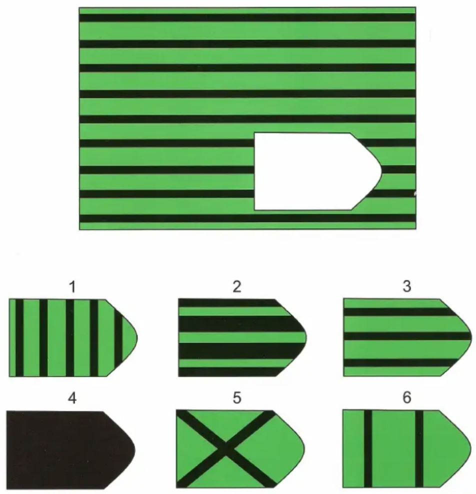
Appendix Table A3: Raven's Colored Progressive Matrices

Outcomes	Definitions
From Colored Progressive Matrices Set A	*Six options per item*

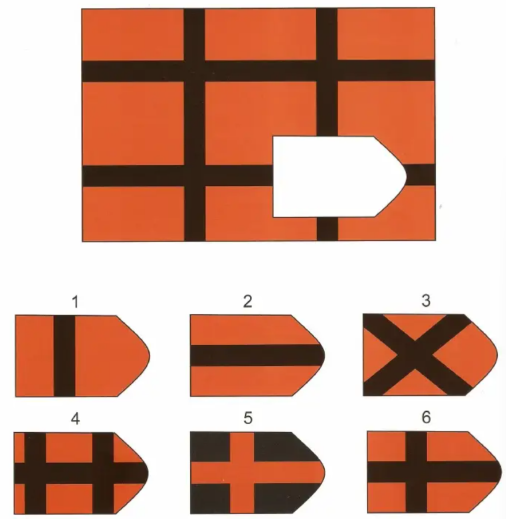




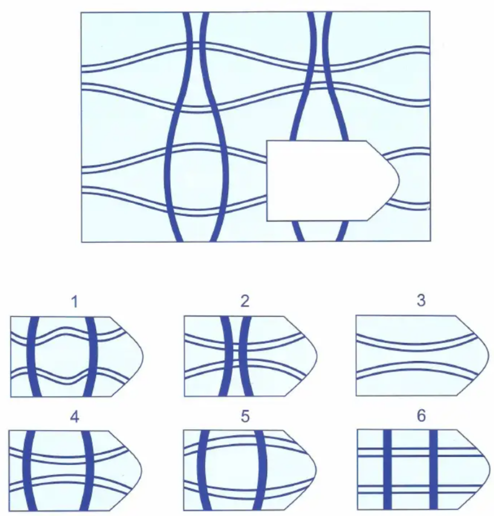
A6



A7



A11



Appendix Table A4: Other Assessment Items

Outcomes	Definitions
ASER Problems	*Open response* 82 _ 64 51 _ 28 8) 984 (
PASEC Read Questions	*Multiple-choice responses* Read the following passage and then answer the questions below: Today we have seen a nurse. She has vaccinated all the pupils and the teacher against COVID/Corona virus. The nurse pricked my arm as hard as a mosquito. 1. Who came to school today? (a) A teacher (b) A Merchant (c) A nurse (d) A mosquito 2. The children were vaccinated against… (a) COVID/Coronavirus (a) Mosquitoes (c) the flu (d) Measles

Appendix Table A5: Student Study Preferences

Outcomes	Definitions
Study Preferences	*Student self-report responses* The questions that follow are about your school-life. 1. Put the school subjects you study in order from your favorite to least favorite. The subject you like the most should come first. (a) Science classes (b) Math classes (c) Hindi classes (d) Social science classes 2. How many classmates do you play with? 3. How many classmates do you study science with? 4. For subjects other than science, how many classmates do you study with?
Education Aspirations Index	*Student self-report responses* 1. If finances were not a problem, and a good school/college was available, what is the highest level of education you would like to complete? (a) Middle School (8th) (b) High School (10th) (c) Higher Secondary (12th) (d) Vocational / Technical School (e) College / University / Bachelors (BA/BSc) (f) Master's (MA/MSc) (g) PhD / Doctorate 2. If you were to go to college, which of the following majors would you be most interested in pursuing? (a) Engineering (b) Medicine (c) Science (d) History (e) Maths (f) Hindi (g) English (h) Other (please specify) (i) I don't know yet 3. If you were to go to college, would you prefer to stay in Fatehpur or move to another city? (a) I would like to stay in Fatehpur for college. (b) I would like to move out for college. (c) I don't know yet 4. In two years' time, how confident are you that you will be enrolled in school? (a) Fully confident (b) Very confident (c) Somewhat confident (d) Not very confident (e) Not at all confident 5. Think about your past exam scores. If you study hard, how many marks out of 100 do you think you can get in your next science exam?

Appendix Table A6: Student Study Preferences

Outcomes	Definitions
Student Time Table	*Responses: Yes/No; minutes* 1. Do you take private tuition? (a) Yes (b) No 2. On a typical school day last week, how much time did you spend doing self-study or homework (not including tuition or school hours)? 3. On a typical school day last week, how much time did you spend studying at home (not tuition) for the following subjects: (a) Science (b) Maths (c) Other Subjects 4. On a typical school day last week, how much time did you spend talking with friends, watching TV, playing video games, or doing other fun activities? 5. On a typical school day last week, how much time did you spend helping your family with household work or other responsibilities (like farming, cooking, cleaning, or caring for younger siblings/elders)?

Appendix Table A7: Student-Reported Classroom Activities

Outcomes	Definitions
Classroom Activity	*Responses: 1 - Never; 2 - Rarely; 3 - Sometimes; 4 - Often* 1. My science teacher does things that make me feel confident in my ability to do well in the course. For example, they tell me that I can do well if I keep trying. 2. When my teacher explains a difficult idea, I understand it better. 3. The science teacher gives extra help when students need it. 4. My teacher has clear answers to my questions. 5. When I don't understand something in my science class, I ask questions to understand it better. 6. In my science class, I ask questions the answers to which are not yet covered by the teacher. 7. I do not hesitate in asking questions in my science classes. 8. When I give an answer, my teacher makes me explain why I think it is correct. 9. I work together with classmates in a small group on an in-class assignment or activity. 10. Class discussions where we are encouraged to ask questions and share our ideas. 11. We repeat facts out-loud together in class to memorize them. $(-)$ 12. I work quietly by myself on an assignment. $(-)$

Appendix Table A8: Teacher Curiosity Index

Outcomes	Definitions
Curiosity Index	*Response Options: 1 - Strongly disagree; 5- Strongly agree* 1. Mysterious things which are hard to understand make me curious. 2. It gets on my nerves if I am close to solving a question but can't figure it out fully. 3. I am close to solving a question but can't figure it out fully. I get ambitious. 4. It is frustrating not knowing a question. So, I try working harder to learn it. 5. Sometimes, when I come across a science idea that's only briefly mentioned, I wonder about it later.

Appendix Table A9: Teacher Science Questions

Outcomes	Definitions
Teacher Science Test	*Multiple-choice responses* 1. How many centimeters are there in a meter? (a) 1000 cm (b) 200 cm (c) 100 cm (d) 10 cm 2. How many parts are there in the human brain? (a) Two (b) Three (c) Four (d) Five 3. The building block of matter is (a) Atom (b) Electron (c) Proton (d) Neutron 4. The motion of a pendulum in a clock is an example of – (i) Rectilinear motion (ii) Oscillatory motion (iii) Rotational motion (iv) Circular motion 5. Which of the following is a chemical change? (a) Bulb glowing (b) Fan running (c) Rusting (d) Ice melting 6. Insectivorous plants are generally found in areas where the soil lacks - (i) Oxygen (ii) Water (iii) Nitrogen (iv) Carbon 7. The boiling point of water is- (a) 10°C (b) 100°C (c) 120°C (d) 40°C 8. Mechanical energy changes to – when two stones strike rapidly (i) Light and sound energy (ii) Heat and sound energy (iii) Heat and light energy (iv) Heat, light, and sound energy 9. The unit of pressure is - (a) Newton/meter (b) Kilogram (c) Joule (d) Newton/m^2 10. Which of the following is a Rabi crop- (a) Rice (b) Corn (c) Peanuts (d) Wheat 11. Which of the following is attracted to a magnet - (a) Sawdust (b) Glass piece (c) Iron filings (d) Copper filings 12. Which particle has a negative charge? (i) Proton (ii) Electron (iii) Neutron (iv) All

Appendix Table A10: Teacher Time Table

Outcomes	Definitions
Teacher Activities	On average, how often do you do the following in this school? *Responses: Never; Once a year or less; 2–4 times a year; 5–10 times a year; 1–3 times a month; Once a week or more* 1. Administrative duties not for your school (including managing elections and other tasks outside of your school mandated by the government). 2. Facilitating activities for your students outside of class time. 3. Communicating with parents or guardians. 4. Exchanging ideas and teaching materials or getting help from other teachers in the same school. 5. Exchanging ideas and teaching materials or getting help from other teachers at a different school. 6. Attend administrative or teacher meetings. 7. Discuss with other teachers about specific students. 8. In a typical period of class time, how many minutes do you spend on average for each activity: (*Total minutes must add up to 40*). a) Lecturing b) Students work on assignments quietly on their own c) Students work on assignments in groups or pairs (not involving hands-on experiments and games d) Interactive or hands-on activities like experiments and games e) Keeping order in the classroom (maintaining discipline) f) Chants or call and response games to ensure students remember facts g) Classroom discussions h) Other activities

Appendix Table A11: Teaching Philosophy

Outcomes	Definitions
Teacher Locus of Control	*Response Options: 1 - Strongly disagree; 5- Strongly agree* 1. When a student gets a better grade than he/she usually does, it is usually because I found better ways of teaching that student. 2. When I really try, I can get through to the most difficult students. 3. I am very limited in what I can achieve because a student's home environment is a large influence on his/her achievement. 4. Teachers are not a very powerful influence on student achievement when all factors are considered. 5. The influences of a student's home experience can be overcome by good teaching. 6. If I don't do my job as a teacher well, my students' future and families' future will suffer. 7. I am enthusiastic about initiating innovations and changes at my school.

	8. I have enough experience to know which teaching methods work best in the classroom. 9. Trying new approaches to teaching is risky, my students will learn better by sticking with the methods I have always used. 10. I feel full of energy and readiness to solve any problems which arise at my school.
Teaching Philosophy	*Response Options: 1 - Strongly disagree; 5- Strongly agree* 1. It is better when the teacher – not the student – decides what activities are to be done. 2. My role as a teacher is to facilitate students' own inquiry. 3. Teachers must lead the classroom because students will not understand the material if they try to learn on their own. 4. Students learn best by finding solutions to problems on their own. 5. How much students learn depends on how much background knowledge they have – that is why devoting class time to memorizing facts is so important. 6. Students should be allowed to think of solutions to practical problems themselves before the teacher shows them how they are solved. 7. Students must follow procedures or directions exactly to ensure they arrive at the correct result or answer. 8. Thinking and reasoning processes are more important than specific curriculum content. 9. Students deserve more of my attention if they are lagging behind in classwork or homework.
Active Learning Practices	**Activities:** 1. Students work together in small groups on an in-class assignment or activity. 2. Class discussion where students are encouraged to ask questions and share their ideas. 3. Students repeat facts out-loud in class to memorize them. 4. Students work quietly by themselves on an assignment. **Each activity description is followed by two questions:** 1. How often does **the activity** happen in your science classes throughout the year? *Response Options: 1 - Never; 2 - A few times a year; 3 - About once a month; 4 - Two or three times a month; 5 - once a week; 6 - Multiple times a week* 2. Whether or not **the activity** happens, do you agree or disagree that **the activity** is an effective method for teaching science? *Response options: 1 - Strongly disagree; 5 - Strongly agree*

Appendix Table A12: Teacher Agency

Outcomes	Definitions
Teacher Agency	*1 - No Discretion; 5 - Complete Discretion* How much discretion do you exercise over the following aspects of your role as a teacher? 1. Textbook, assignments, and other learning materials 2. How to spend your time at work, outside of class time 3. The order in which to cover different parts of the curriculum 4. Whether to skip or only briefly cover a section of the curriculum 5. Final grades at the end of the school year 6. Discipline over student misbehaviour 7. Teaching style or pedagogy 8. Including activities that go slightly beyond the textbook, as long as students understand the curriculum.

Appendix Table A13: Basic Information and Professional Development

Outcomes	Definitions
Basic Information	1. What is your full name? 1.1 Phone number 1.2 Gender 1.3 Age 2. What is your position as teacher? a) Head Teacher b) Acting Head Teacher c) Assistant Teacher d) Instructor (as per RTE) 3. When did you join this school as a teacher? 4. What is the total number of years you have taught in public schools? 5. What is the total number of years you have taught in private schools? 6. During your formal training to become a teacher? What area did you specialize in? (Select all that apply) *Response Options: Maths, English, Hindi, Other Language, Science, Social Science, Other* 7. What is the main subject you teach? *Response Options: Maths, English, Hindi, Other Language, Science, Social Science, Other* 8. How well do you speak English? *Response Options: 1 - No English at all; 2 - A little; 3 - I am fluent.* 9. In which of these grades do you teach science? (Select all that apply) *Response Options: Grade 6, Grade 7, Grade 8*
Professional Development	1. Do you believe that other types of teacher training programs can improve your effectiveness as a teacher? *Response Options: 1=Yes; 2=No.*

	2. Have you ever attended a professional development training for improving your teaching abilities? *Response Options: 1=Yes; 2=No (If No, skip to 2.3).* 2.1 In all, how many days of professional development have you attended in the last year? 2.2 How would you rate the match between what you learned in training and what you need to teach effectively in the classroom? Very little match – Training and classroom needs are quite different. Some match – A few things from training apply to my classroom. Moderate match – About half of what I learned is useful. Good match – Most training content fits my needs. Excellent match – Training closely aligns with what I need. 2.3 Since you started teaching, have you participated in any of the following kinds of professional development activities? *Response Options: 1=Yes; 2=No.* Courses/workshops (e.g., on subject matter, methods, or education topics) Qualification programme (e.g., a degree programme) Participation in a teacher network for professional development Mentoring or peer observation as part of a formal school arrangement

Appendix Table A14: Teacher Knowledge of Students

Outcomes	Definitions
Teacher Knowledge of Students	*Integer responses* 1. How many of your students have a household member who lives outside of Fatehpur? 2. How many of your students have at least one parent who completed upper primary school? 3. How many of your students have at least one parent who completed secondary school?

Appendix Table A15: Social Desirability Scale

Outcomes	Definitions
Social Desirability Index	*Responses: Agree; Disagree*
	1. It is sometimes hard for me to go on with my work if I am not encouraged.
	2. I sometimes feel resentful when I don't get my way.
	3. On a few occasions, I have given up doing something because I thought too little of my ability.
	4. There have been times when I felt like rebelling against people in authority even though I knew they were right.
	5. No matter who I'm talking to, I'm always a good listener. $(-)$
	6. There have been occasions when I took advantage of someone.
	7. I'm always willing to admit it when I make a mistake. $(-)$
	8. I sometimes try to get even rather than forgive and forget.
	9. I am always courteous, even to people who are disagreeable. $(-)$
	10. I have never been irked when people expressed ideas very different from my own. $(-)$
	11. There have been times when I was quite jealous of the good fortune of others.
	12. I am sometimes irritated by people who ask favors of me.
	13. I have deliberately said something that hurt someone's feelings.

Appendix Table A16: Principal Survey Instrument

Outcomes	Definitions
Basic Information	1. What is your full name? 1.1 Phone number 1.2 Gender 1.3 Age 2. In what year did you join this school? 3. In what year did you become a principal at this school? 4. Have you ever been a teacher at this school? *1-Yes; 0-No; If No, skip to 6* 5. If yes, what was the last year you taught? 6. How well do you speak English? *Response options: 1-No English at all; 2-A little; 3-I am fluent* 7. During the last school year (2024–25), has the school received any support from organizations such as NGOs, faith-based organizations, private individuals, or companies? *Response options: 0-No (skip to 10); 1-Yes* 8. How much funding in rupees did such organizations provide during the last school year (2024–25)? 9. Did such organizations provide any services or non-monetary donations during the last school year (2024–25)? *Response options: 0-No; 1-Yes*
Evaluation as Principal	1. As a principal, you must respond to many stakeholders who may formally or informally evaluate your job performance. Which of the following stakeholders are most important for determining your job security? *Response Options: 1 - Not at all important; 2-Only a little important; 3-Important; 4-Very important* a. Central government officials b. State-level government officials c. District-level government officials d. Non-teaching school staff e. Teachers f. Students g. Parents h. Non-government community leaders i. Non-profit or faith-based organizations 2. Are there any other stakeholders that officially evaluate your performance as a principal at this school? *Response options: 1-Yes; 2-No (If no, skip to next section)* 2.1 Please list any other stakeholders who may evaluate your performance as a principal at this school.
Teachers' Performance	*Instruction: Please select the five most important criteria for evaluating a teacher's performance.* 1. Subject Matter Expertise 2. Lesson Planning and Delivery 3. Classroom Management

	4. Student Learning Outcomes 5. Student Engagement and Motivation 6. Assessment and Grading 7. Attendance and Punctuality 8. Support during the Absence of Other Teachers 9. Communication with Parents and Guardians 10. Professional Development and Growth 11. Teacher has covered the entirety of the set curriculum 12. Are there any other criteria your school uses to evaluate teacher performance? *Response options: 1-Yes; 2-No* (*If no, skip to next section*) 12.1 Other (Please Specify)
Teacher Monitoring and Evaluation	1. When a teacher is absent or on leave, how often does the school adopt the following strategies? *Response Options: 1-Never; 2-Rarely; 3-Sometimes; 4-Often* a. Someone who does not commonly teach at this school teaches the lesson. b. Headmaster/principal teaches the lesson. c. A teacher at this school teaches the lesson, without combining students from different grades. d. A teacher at this school teaches the lesson combining students from different grades. e. Students are assigned to work on academic tasks, such as reading or homework. f. Students are assigned to work on non-academic tasks, such as gardening, exercise, or play. g. Students of the class are allowed to go home. 2. Consider the following approaches to monitoring teacher performance. 2.1. Do you use this approach to evaluate teacher performance? *Response Options: 1-Yes; 2-No* 2.2. If Yes, then: How often does your school use this method? *Response Options: 1-Never; 2-Rarely; 3-Sometimes; 4-Often* a. Unannounced inspection or classroom observation by principal or school staff. b. Inspection by external observer, such as a government official. c. Student complaints. d. Complaints by teachers and other staff. e. Reviewing teacher records, including attendance and student marks. f. Other (Specify). 3. What is the first course of action the school would take with teachers who are not performing well? *Response options: 1-Issue a warning; 2-Complain to higher officials; 3-Meeting with teacher; 4-Other* (*specify*)
Changes in Education Policy	How much do you agree with the following statements and opinions about your experiences as a principal in this school? *Response Options: 1-Strongly disagree; 2-Disagree; 3-Neither agree nor disagree; 4-Agree; 5-Strongly agree*

	1. I am enthusiastic about initiating innovations and changes at my school. 2. When changes in education policy happen, I feel I can handle them easily. 3. Changes in education policy ultimately improve the quality of my school. 4. I feel full of energy and readiness to solve any problems which arise at my school. 5. Changes in education policy do not come with promised support. 6. Education policy does not change even with a change in government. 7. Each government brings changes to education policy, but these changes do not have an impact on my students.
Past Teacher Training Programs	Instruction: If your school has never participated in an in-service teacher training program, please skip this section. 1. How often do teachers in your school attend in-service training programs? *Response options: 0-At least once a month; 1-Every 6 months; 2-Once a year; 3-Every two years; 4-Our school does not regularly participate.* 2. When was the last time teachers attended an in-service training program? *Response options: 1-In the last 6 months; 2-In the last 2 years; 3-More than 2 years ago (If more than two years, skip to next section).* 3. Have teachers received training in subject material in the last two years? *Response options: 0-No; 1-Yes.* 4. Have teachers received training in the use of teaching materials in the last two years? *Response options: 0-No; 1-Yes.* 5. Have teachers received training in any of the following teaching styles in the last two years? (Select all that apply) 00-Teacher-led classroom 01-Inquiry-based learning 02-Rote learning 03-Lecture-based classroom 04-Active learning 05-Discussion-based classroom 06-Constructivism 07-Passive learning 08-Hands-on/experiential learning 09-Other (please specify) 6. Which organizations or programs have provided teacher training services in the last two years? (Select all that apply) 01-External experts/consultants 02-Government programs 03-Non-government organization (NGO) 04-Other (please specify)

	7. When teachers do not implement what they learned from training, what are the three most common reasons? (Select up to three options) 01-Lack of time 02-Lack of resources/materials 03-Lack of support 04-Program was of poor quality 05-Resistance to change 06-Other (please specify) 8. What specific improvements in student learning or marks have been observed as a result of teacher training in the last two years?
Hypothetical Scenarios and Training Needs	Instruction: The following questions are hypothetical. Your answers will not affect your school's ability to participate in any program. 1. Imagine your school was invited to participate in a new training program. Would you support the participation of your teachers? *Response options: 1-Yes, definitely; 2-Yes, if aligned with school goals; 3-No, not at this time; 4-Not sure.* 2. Would you ensure their attendance in the training? *Response options: 1-Yes, monitor attendance; 2-Yes, encourage but not monitor; 3-No, leave to teacher discretion; 4-No, ask not to attend.* 3. How would you encourage their attendance? (Select all that apply) 01-Providing incentives or rewards 02-Adjusting the school schedule 03-Emphasizing the importance of the training 04-Holding discussions or meetings to address concerns 05-Other (please specify) 4. Please indicate the extent to which you believe teachers at your school need professional development or training in the following areas. *Response options: 1-Don't need additional training; 2-Need some training; 3-Need a lot more training.* a. Their knowledge of the curriculum and standards in science education b. How to effectively assess student achievement c. Instructional/pedagogical practices d. Effective use of educational materials/resources (textbooks, technology, etc.) e. Managing the classroom, student behavior, and discipline f. Student career guidance and counseling g. Helping students develop cross-subject skills such as problem-solving and learning how to learn h. How to effectively communicate and collaborate with parents

A.2 Steps for Index Construction

Our study employs variance-weighted indices constructed following the methodology proposed by Anderson (2008).[14] For implementation, we utilize the Stata module SWINDEX developed by Schwab et al. (2021), which facilitates the calculation of these indices.

Many of our outcome variables comprise multiple individual items, typically measured on five-point Likert scales or agree-disagree scales. To create composite indices from these items, we aggregate them using a weighted-average approach. The weighting scheme normalizes individual items to share a common standard deviation, with weights derived from the inverse covariance matrix as specified in Anderson (2008). This procedure ensures that our indices appropriately account for the correlation structure among constituent items while giving greater weight to those with lower covariance with other components.

Below we outline the steps we follow for index construction based on Anderson (2008).

1. All variables are recoded such that a higher score consistently represents a more favorable outcome. If a question's original scale was reversed, this is corrected so that, after recoding, higher values always denote a positive result.

2. For questions based on a 5-point Likert scale, we also generate a corresponding binary variable. This binary indicator is assigned a value of one if the participant selected "Strongly Agree" or "Agree" (for positively-phrased statements) or "Strongly Disagree" or "Disagree" (for negatively-phrased statements), and zero otherwise.

3. Next, each variable is standardized by subtracting its mean and dividing by the standard deviation observed in the control group. Call this the standardized variable ($\tilde{y}$).

4. We then construct the covariance matrix ($\widehat{\sum}$) for the set of standardized variables. Each element of this matrix is calculated as the product of the centered and standardized scores for each pair of variables, summed over all available observations.

$$\sum_{mn}^{\wedge} = \sum_{i=1}^{Nmn} \frac{(y_{im} - \bar{y}_m)}{\sigma_m^y} * \frac{(y_{in} - \bar{y}_n)}{\sigma_n^y}$$

[14] Refer to Haushofer and Shapiro (2016); Fiala et al. (2022) for a recent application.

where, N_{mn} is the number of observations (total persons with non-missing data for variables m and n).

5. The covariance matrix is inverted, and for each variable, its corresponding weight (w_k) is calculated as the sum of entries in the respective row of the inverted matrix.

$$(\hat{\sum})^{-1} = \begin{bmatrix} c_{11}....c_{1K} \\ .... \\ .... \\ c_{K1}....c_{KK} \end{bmatrix}$$

$$w_k = \sum_{l=1}^{K} c_{kl}$$

6. In the final step, we construct the composite index $\hat{y}_i$ for each individual by computing a weighted average of their standardized scores, $\tilde{y}_{ik}$.

$$\hat{y}_i = \Big(\sum_{k \in K} w_k\Big)^{-1} \sum_{k \in K_i} w_k * \frac{y_{ik} - \bar{y_k}}{\sigma_k^y}$$

A.3 Field Operations

A.3.1 Field Team Composition

The field team from DAI Research and Advisory Services consisted of 80 enumerators, 10 field supervisors, and 2 project assistants. Enumerators were organized into ten sub-teams of eight members each to ensure efficient coordination and monitoring. Each sub-team was managed by one field supervisor, while the two project assistants jointly oversaw all sub-teams and served as the main liaison between field staff and the central research team. The entire field operation reported to one Research Associate, who was responsible for quality assurance and for ensuring that data collection and intervention activities adhered to the research design, ethical protocols, and daily operational standards.

A.3.2 Survey Data Collection and Quality Control

Baseline data collection was conducted during July and August 2025 at the student, teacher, and principal levels. Prior to fieldwork, the team obtained all necessary approvals from school management and collected written consent from student–parent pairs, teachers, and principals. Surveys were administered during pre-scheduled school visits by the field teams.

Student surveys were conducted in classroom settings using tablets pre-loaded with the digital questionnaire designed in SurveyCTO. The survey was self-administered but guided by enumerators, with an approximate ratio of one enumerator for every six students to provide clarification on technical or conceptual issues. Enumerators received intensive training to ensure standardized administration, resolve student queries, and maintain a supportive environment throughout the process.

Teacher and principal surveys were also conducted digitally but were fully self-administered and unsupervised. Teachers and principals were given detailed instructions and contact information for assistance in case of any questions; however, enumerators did not remain present for the entire duration of these surveys to minimize response bias. All responses were synced daily to the SurveyCTO server and reviewed by the Research Associate to verify completeness, accuracy, and adherence to data quality protocols.

A.3.3 Field Notes

Student Attendance in Fatehpur Schools

Preliminary observations from the baseline data indicate that student attendance in government upper-primary schools across Fatehpur is considerably lower than expected. The average attendance rate during our visits was 49.3%, with substantial variation across blocks. In several schools, fewer than half of enrolled students were present on a typical day.

Low attendance poses both logistical and pedagogical challenges for the intervention. On the one hand, it reduces the effective reach of curiosity-based classroom activities; on the other, it highlights deeper structural constraints facing students and their families. In particular, we identified three main factors contributing to persistently low attendance:

1. **Enrollment inflation under RTE:** Under the Right to Education Act, teachers and headmasters are assigned enrollment targets that they are expected to meet. This often results in inflated registers, with many students recorded as enrolled but seldom attending classes in practice.

2. **Agricultural labor:** Many children are withdrawn from school during peak agricultural seasons to support household labor requirements, leading to irregular attendance patterns.

3. **Private schools:** A substantial proportion of students are nominally enrolled in government schools but actually attend nearby low-cost private schools. Their names remain on official registers, but their participation in government classrooms is effectively absent.

Documenting this low baseline attendance is critical for interpreting program impacts. Because the study examines how curiosity-driven pedagogy influences engagement and learning outcomes, reduced classroom participation may attenuate measured effects. This pattern also aligns with broader concerns about chronic absenteeism in rural Uttar Pradesh. According to UDISE+ data for neighboring districts, average student attendance in upper-primary grades ranges between 65–72%, reflecting similar systemic challenges.[15] These findings suggest that pedagogical innovations may need to be coupled with complementary strategies to improve regular attendance and ensure sustained student engagement.

[15] Authors' calculations from UDISE+ 2022–23 data for Fatehpur, Banda, and Kaushambi districts.

Survey Implementation Timeline

The pilot of the survey instrument was conducted in two phases, from June 15 to June 19, 2025, and again from July 1 to July 5, 2025. The break in the middle coincided with extended summer holidays in Fatehpur schools. Prior to data collection, education officers in each block were contacted to request their cooperation. The block officers kindly agreed to convey official permission for our survey to all sample schools via their administrative WhatsApp groups.

Field operations proceeded largely as planned, with a few notable events:

1. **July 22, 2025:** One school was replaced after a teacher declined participation, to ensure smooth field implementation.

2. **August 5, 2025:** Survey operations were temporarily halted for two days due to heavy rainfall and localized flooding.

3. **August 10, 2025:** Additional teacher surveys were conducted following the Rakshabandhan break.[16]

[16] Rakshabandhan is a traditional Hindu festival that celebrates the bond of love, protection, and duty between brothers and sisters.

A.4 Curriculum Details by Grade

This appendix presents the detailed structure of the ten sessions used in delivering the *Curiosity Pedagogy*, jointly developed by the research team and the Agastya International Foundation. The curriculum was adapted from Agastya's prior science education modules and aligned with the upper-primary syllabus of Uttar Pradesh. Each session was pilot-tested in two government schools in Fatehpur to ensure age appropriateness, material feasibility, and smooth integration with regular classroom schedules. Activities were refined based on teacher feedback and student engagement observations during the pilot phase. All sessions were designed for a 45–60 minute classroom period using locally available, low-cost materials such as soil, candles, bottles, turmeric, nails, and magnets.

The overarching pedagogical framework emphasizes learning by doing and thinking by questioning. Each lesson follows a structured sequence beginning with a "super start"—a surprising demonstration or guiding question designed to elicit curiosity—followed by hands-on experimentation, group reasoning, and reflection. This design connects textbook concepts to students' everyday experiences in agriculture, household life, and the natural environment, with the dual objective of improving scientific understanding and cultivating curiosity, reasoning, and confidence in expressing ideas.

A.4.1 Curriculum for Grade 6

1. **Food and Nutrition (Sessions 1–2).** Students explore why humans need diverse foods and test for starch, protein, and fat using simple reagents. Through "food-detective" activities, they learn to read labels, link nutrients to body functions, and recognize healthy dietary habits.

2. **Light and the Pinhole Camera (Sessions 3–4).** Learners build and modify simple pinhole cameras to observe how light travels in straight lines and produces inverted images. Activities highlight image clarity, aperture size, and everyday analogies—from tree-leaf shadows to early cameras—illustrating how scientific observation explains familiar phenomena.

3. **Transpiration and Capillarity (Sessions 5–6).** Using potted plants, colored water, and paper strips, students visualize water movement in plants. They connect transpiration and capillary action to evaporation in nature, earthen pots, oil lamps, and the water cycle, developing appreciation for hidden processes in living systems.

4. **Separation of Mixtures (Sessions 7–8).** Through filtration, magnetism, chromatography, and a hand-spun centrifuge, students discover how physical properties such as particle size, solubility, and density determine separation methods. They relate these principles to local issues such as clean-water access and pollution detection.

5. **Water Cycle and Conservation (Sessions 9–10).** Learners model the global distribution of water, quantify limited freshwater resources, and act out stages of the water cycle. Subsequent experiments and role-plays link human activity, pollution, and natural filtration by soil and plants, reinforcing stewardship of shared resources.

A.4.2 Curriculum for Grade 7

1. **Acids, Bases, and Indicators (Sessions 1–2).** Students identify acids and bases using litmus paper and then create their own turmeric indicator to test household substances. They learn to observe, record, and explain color changes, developing both vocabulary and confidence in scientific reasoning.

2. **Transfer of Heat (Sessions 3–4).** Through experiments on *conduction* and *convection*, students visualize how heat moves through solids, liquids, and gases. Activities with metal rods, wax, and colored water demonstrate particle vibration and fluid movement, linking science to everyday experiences such as cooking and ventilation.

3. **Lung Capacity and Respiration (Sessions 5–6).** Learners measure chest expansion and air volume using simple tools such as measuring tape and a water-displacement setup. These sessions connect physiology to health and exercise, showing how breathing relates to energy, fitness, and observation of body processes.

4. **Buoyancy and Floating (Sessions 7–8).** Using buckets, stones, and spring balances, students explore why objects feel lighter in water and how liquids exert upward forces. They compare buoyant forces across liquids such as water and oil, discovering how density influences floating and sinking.

5. **Physical and Chemical Change (Sessions 9–10).** Students distinguish between reversible physical changes and irreversible chemical ones. Experiments with wax melting versus wood burning introduce evidence of new substance formation. Later, controlled group experiments

with baking soda, vinegar, and water teach experimental design—using controls, isolating variables, and classifying changes through observation.

A.4.3 Curriculum for Grade 8

1. **Soil and Its Properties (Sessions 1–2).** Students investigate the composition and layers of soil through sedimentation experiments and learn to measure water absorption using improvised funnels and soil samples collected from their surroundings. Activities link soil texture to agriculture, pottery, and construction, emphasizing how physical properties shape local livelihoods.

2. **Chemical Reactions and Energy (Sessions 3–4).** Through exothermic and endothermic demonstrations—mixing quicklime with water and combining baking soda with vinegar—students observe how heat can be released or absorbed during reactions. They measure temperature changes, interpret data, and relate findings to real-world processes such as heating packs, cooling effects, and combustion.

3. **Air and Water Pressure (Sessions 5–6).** Learners explore invisible forces through simple experiments with inverted tumblers and perforated water bottles, revealing that both air and water exert pressure. These demonstrations help explain common phenomena such as pumping groundwater, water supply systems, and kite flying, connecting textbook concepts to rural contexts.

4. **Magnetism (Sessions 7–8).** Students make temporary magnets by rubbing an iron nail with a fridge magnet, observe alignment with Earth's poles, and test how heat or impact weakens magnetism. The sessions build understanding of magnetic forces, storage practices, and everyday applications—from fridge doors to local craftsmanship.

5. **Structure of the Atom (Sessions 9–10).** Using paper-folding and scaling activities, students grasp the relative size of atoms and learn to use powers of ten. They then construct atomic models with colored buttons to represent protons, neutrons, and electrons, introducing concepts of atomic number, mass, electron shells, and valency through visual modeling and collaborative reasoning.

Across all grades, the curriculum emphasizes *learning by doing* and cultivating *scientific habits of mind*—observation, prediction, measurement, and explanation. Teachers and facilitators are encouraged to draw on students' everyday experiences in farming, household work, and community life, transforming familiar practices into gateways for conceptual understanding. Each activity is deliberately low-cost, context-sensitive, and inclusive, ensuring that curiosity-driven pedagogy can be implemented effectively in resource-constrained rural classrooms.

A.5 Facilitator Survey

Appendix Table A17: Facilitator Questionnaire

Question	Response Options / Values
Demography	
1. Name	
2. Age (in years)	
3. Highest level of education completed	
4. Marital status	
5. Number of children	
6. Social category	*SC / ST / OBC / General*
7. Place of birth	
7.1 Sub-caste or Jati	
8. Current place of residence	
8.1 State	
8.2 District	
8.3 Village / City	
9. Gender	*Male / Female / Other*
Experience	
10. Have you ever worked as a facilitator before?	*Yes / No*
10.1 If yes, for how long? (in years/months)	
11. Have you ever attended a course on socio-emotional learning (SEL)?	*Yes / No / Don't remember*
12. Have you attended any science-based training before?	*Yes / No / Don't remember*
13. Have you ever worked with children before?	*Yes / No / Don't remember*
Cantril Ladder (Life Satisfaction)	*Scale 1–10*
14. On which step of the ladder do you personally feel you stand at this time?	
15. On which step do you think you will stand about five years from now?	
Behavior and Preferences	*1 - Strongly Agree; 5 - Strongly Disagree*
16. You enjoy handling problems that are completely new to you.	
17. You try to help people understand the underlying concepts behind the point you are making.	
18. You consider cultural or social barriers when planning your sessions.	

A.6 Monitoring Activities

To continuously monitor implementation across both treatment arms, we employed multiple complementary monitoring approaches throughout the intervention period. These approaches are described below and organized by treatment arm.

A.6.1 T1: Teacher Training Arm

The teacher training arm focused on increasing curiosity among both teachers and students. Monitoring therefore began with the teacher training sessions. To measure teacher take-up and participation, we collected the following information during each of the three training rounds.

Teacher-level attendance. Attendance was recorded at the individual teacher level for each training session. Since some schools had more than one participating science teacher, attendance was not aggregated at the school level. Based on these records, make-up training sessions were scheduled for teachers who were absent.

Participation during training. During all training rounds, we used a structured classroom observation tool to capture teacher engagement and participation during the sessions. Following the completion of the training sessions, teachers had full discretion to schedule and conduct experiments in their classrooms. This made direct observation of classroom implementation infeasible. To address this limitation, we conducted two rounds of unannounced midline classroom observations, which served as random checks for T1 schools. Details of the midline observation instrument are provided later in this appendix.

In addition, the following monitoring activities were implemented during the intervention period.

Spot checks. During the initial weeks following the first training round, the field team collected qualitative information from headmasters, teachers, and students to verify whether laboratory materials intended for T1 schools had been received and were being used. The objective of these checks was to confirm that implementation had commenced in each school. Spot checks were discontinued after the first two weeks.

WhatsApp documentation. WhatsApp groups were used as a lightweight documentation mechanism during implementation. Teachers shared photographs and short videos of classroom activities as they were conducted. This media was compiled into a qualitative dataset documenting whether and how specific experiments were implemented, as well as observable aspects of teacher participation and student engagement.

Biweekly calls. Fortnightly review calls were conducted at the block level with participation from teachers, headmasters, and block-level officers. Prior to each call, teachers completed tracking sheets documenting activities implemented during the preceding period. During the calls, teachers reported progress by referring to these sheets, facilitating regular monitoring and coordination across schools within each block.

A.6.2 T2: External Instructor Arm

Monitoring was both more critical and more feasible in the instructor-led treatment arm. Through our implementation partner, Ecoprism, we obtained access to weekly schedules for all facilitators, enabling the DAI data collection team to conduct random spot checks. The following monitoring data were collected in T2 schools.

Spot checks. Based on facilitator schedules shared by Ecoprism, the DAI field team conducted spot checks in T2 schools using an adapted version of the Stallings classroom observation method. Enumerators recorded observations at five-minute intervals using a structured questionnaire.

These spot checks provided information on program fidelity along two margins. At the extensive margin, they verified facilitator attendance and whether scheduled sessions were conducted. At the intensive margin, they captured detailed information on facilitator performance, student engagement with experiments, and use of laboratory materials.

Student session attendance. For each session led by external facilitators, student attendance was recorded by the facilitator either at the beginning or end of the session. This measure reflects attendance at intervention sessions rather than overall school attendance.

Facilitator self-reported monitoring. Ecoprism implemented a self-reported digital monitoring tool completed by facilitators. This instrument included questions on session timing, student interactions, and other implementation details. These data were shared with the field team throughout the intervention period and used to corroborate findings from spot checks.

A.6.3 Midline Classroom Observations

Midline data were collected in two rounds through random, unannounced visits to all schools in the sample, including T1, T2, and control schools. District officials informed science teachers that such visits would occur, but teachers were not informed of the specific dates.

Because the midline observations were conducted during regular science classes across all treatment arms, they also functioned as a monitoring tool for T1 implementation. The full midline classroom observation protocol and instrument are provided below.

Protocol for Midline Observation:

1. This is an observation form that needs to be filled by the enumerator upon observing science classes in the schools.
2. Enumerators must arrive at the school at least 15 minutes before the scheduled start of the science class. They should observe the entire class from the beginning to the end.
3. If the designated science teacher is unavailable, and another teacher is substituting, the observation should still be completed. The observation should be completed for the designated class hour irrespective of which teacher is teaching which subject.
4. The teacher's consent should be duly obtained before starting the observation.
5. No students or teachers need to be surveyed or interviewed for the collection of this data.
6. The enumerators should not disrupt the classroom activity, or interact with the students while the class is in progress.
7. The only situations in which an observation may be incomplete are when the teacher leading the class does not give consent, or when no teacher is present, i.e. neither the science teacher nor any other teacher is teaching the class.

Appendix Table A18: Midline Survey

Outcome	Response Options / Values
1. Date	
2. Surveyor Name	
3. Block	

4. School Name	
5. Time of Arrival at school	(*HH:MM*)
6. Scheduled class start time	(*HH:MM*)
7. Scheduled class end time	(*HH:MM*)
8. Class	*6th, 7th, 8th*

Teacher Availability and Consent

9. Is the science teacher present at the scheduled time?	*Yes / No*
If 9 = No	
9.1 Reported reason for absence	*Reason unknown* *Sick or medical leave* *Other authorised leave* *Administrative duties* *Other teaching duties* *Science class not happening as per schedule* *Other (specified)*
9.2 Is another teacher substituting for the science teacher?	*Yes/No*
If 9.2 = Yes: Name of the substitute teacher	________________
If 9.2 = No: There is no teacher present in the class. What are the students doing?	*Students are doing something outside the classroom* *Students are sitting in the classroom or studying on their own* *Students are eating* *Students are not present in the school* *Other, specify*
10. Did the teacher consent to the classroom observation?	*Yes/No*
10.1 Reason for refusal, if any	________________
11. Subject being taught during the science period	*Science / Other (specified)*

Classroom Resources and Class Start Details

12. Which of the following items are available in the class? Select all that apply.	*Blackboard / Whiteboard* *Chalk / Markers* *Textbooks for teacher* *Textbooks for students* *Charts / Posters* *Toys / Games* *Science or mathematics equipment* *Laptop / Computer* *Smartboard / LCD Projector / Television* *None of the above*
13. Actual class start time	(*HH:MM*)
14. Number of boys present at the start of the class	

15. Number of girls present at the start of the class

Snapshot Observations - Repeated Every Five Minutes

Answer the questions in this section every five minutes, five times in total. Record what is happening in the classroom **at each moment**.

Teacher Activity

16. What is the teacher doing at the moment?	*Teacher not in the classroom* *Teacher not engaged with students* *Managing discipline* *Reading aloud* *Asking students to read aloud* *Rote learning activities* *Assigning or explaining tasks* *Explaining or lecturing* *Demonstrating or conducting experiments* *Conducting a game or activity* *Other (specified)*
17. What materials are being used by the teacher at the moment?	*None* *Blackboard / Whiteboard* *Textbooks* *Charts / Posters* *Science kit or lab materials* *ICT tools* *Other (specified)*
18. What materials are being used by the students at the moment?	*None* *Blackboard / Whiteboard* *Textbooks* *Notebooks or writing materials* *Charts / Posters* *Science kit or lab materials* *ICT tools* *Other (specified)*

Periodic Observations - Repeated Every Five Minutes

Observe the class for 5 minutes. Answer the questions in this section keeping in mind **the entire span** of five minutes.

Student Behaviour and Interaction

19. Share of students paying attention to the teacher (Listening, watching, or following along)	1. *Almost no one* 2. *Less than half the students* 3. *Almost half of the students* 4. *More than half the students* 5. *Almost all the students*
20. Share of students actively interacting with the teacher (Asking or answering questions, participating in discussion, or performing the task with the teacher)	1. *Almost no one* 2. *Less than half the students* 3. *Almost half of the students*

	4. *More than half the students* 5. *Almost all the students*
21. Did any student ask questions related to the lesson content?	*Yes / No*
22. Did any student ask questions beyond the lesson content?	*Yes / No*

Teacher Behaviour and Interaction

23. How often did the teacher move around in the classroom?	1. *Stayed in one spot for the entire 5 minutes* 2. *Moved around the classroom occasionally* 3. *Moved around the classroom mostly* 4. *Moved around for the entire 5 minutes*
24. What type of questions did the teacher ask? (Closed: Yes/No or one-word answers) (Open-ended: Long answers which require some explanation)	1. *Teacher did not ask questions* 2. *Mostly closed* 3. *Equal mix of closed and open* 4. *Mostly open-ended*
25. Did the teacher do the following things in the 5 minutes?	
1. Connect the lesson to the student's home life	*Yes / No*
2. Introduce real-world phenomenon that sparked questions	*Yes / No*
3. Encourage the students to justify or explain their answers	*Yes / No*
4. Praise or encourage students after answers	*Yes / No*
26. Did the teacher discuss the student's mistakes or incorrect answers?	*Yes / No / No student made a mistake*

End of Class Observations

Class End Details

27. Actual class end time	(*HH:MM*)
28. Boys present at end	(*HH:MM*)
29. Girls present at end	(*HH:MM*)

Overall Observations

30. Did the teacher allow students to ask questions?	*Yes / No*
31. Did the teacher assign independent work to students?	*Yes / No*
Not Homework: For instance, solving a question on their own; looking for answers in a textbook etc.	*Yes / No*
32. Did the teacher engage with students in all parts of the classroom equally, instead of focusing only on one area (like the front or the back of the classroom)?	*Yes / No*