Discussion Paper Series

IZA DP No. 18677

May 2026

Science on the Move: How Experiential Pedagogy Shapes Human Capital

Nitin Kumar Bharti
University of Western Australia

Samreen Malik
New York University Abu Dhabi
and IZA@LISER

Abhiroop Mukhopadhyay
Indian Statistical Institute
and IZA@LISER

Nishith Prakash
Northeastern University, BREAD, CESifo,
IZA@LISER, GLO, HiCN and CReAM

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18677 May 2026

Science on the Move: How Experiential Pedagogy Shapes Human Capital*

Abstract

Despite near-universal school enrollment across many developing economies, the provision of quality education that cultivates lifelong learning and the capacity to apply knowledge in novel circumstances remains elusive. We conduct a cluster-randomized controlled trial in 132 public schools in Uttar Pradesh, India, to evaluate a guided, discovery-based science pedagogy at two intensity levels: a high-intensity Mobile Science Lab (MSL) and a lower-intensity Lab on Bike (LoB). MSL improves motivational beliefs and self-confidence by 0.15–0.18 standard deviations, reduces perceived barriers to education by 0.23 standard deviations, raises engagement by 0.17–0.22 standard deviations, and increases standardized test scores by 0.22–0.34 standard deviations across all subjects. LoB produces limited average effects, with gains concentrated among students completing all sessions. These findings demonstrate that pedagogical design and delivery intensity are critical determinants of multidimensional human capital formation, and that discovery-based pedagogy can shift motivational beliefs, engagement, and achievement in low-capacity public school systems.

JEL classification

C93, D83, I21, I24, O15

Keywords

experiential pedagogy, curiosity, student engagement, randomized controlled trial, human capital, India

Corresponding author

Nishith Prakash

n.prakash@northeastern.edu

* We thank Shubhro Bhattacharya, Palaash Bhargava, Prashant Bharadwaj, Santiago Caicedo, James Dana, Jishnu Das, Alex Eble, David Evans, Deon Filmer, Alejandro Ganimian, Garance Genicot, Rachel Glennerster, Joshua Goodman, Shareen Joshi, Eeshani Kandpal, Shantanu Khanna, Cynthia Kinnan, Dilip Mookherjee, Priya Mukherjee, Robert K. Triest, and participants at seminars in Ashoka University, Bennett University, Center for Global Development, Clark University, Delhi School of Economics, Georgetown University, Higher School of Economics, Indian Institute of Technology–Mumbai, Northeastern University, Patliputra University, Plaksha University, Shiv Nadar University, Teachers College at Columbia University, Thammasat University, Tufts University, University of Namur, Warwick University, World Bank (South Asia Gender Innovation Lab) and conference participants at North East Universities Development Consortium Conference (2024), 18th Annual Conference on Economic Growth and Development (Indian Statistical Institute, Delhi), 7th International Applied Development Economics Conference, and World Congress (2025) for helpful feedback. We thank Dibyajyoti Basak, Rajdeep Chaudhuri, Manvendra Singh, and Bharat Singhal for excellent research assistance. We are deeply grateful to Rama Prasad "Ramji" Raghavan, founder and chairperson of Agastya International Foundation, and to Venkateswar Rao Ketineni, Hariharan Ganesan, Sai Chandrasekhar, and Mukesh Kumar for their invaluable support, guidance, and commitment throughout the implementation of this intervention. We are grateful to the team at Quest Research and Development for their tireless work in carrying out the data collection. This study was approved by the IRB at New York University (IRB No.: HRPP-2022-95) and pre-registered with the AEA Trial Registry (RCT ID: AEARCTR-0010902). We gratefully acknowledge research funding from the New York University Abu Dhabi Development Group, Agastya International Foundation, and the J-PAL Learning for All Initiative (LAI) for conducting fieldwork.

"For the mind does not require filling like a bottle, but rather, like wood, it only requires kindling to create in it an impulse to think independently and an ardent desire for the truth."

— Plutarch, *Moralia*, "On Listening to Lectures," §18 (48C–D),

trans. F.C. Babbitt (Loeb Classical Library, Vol. I, 1927)

1 Introduction

The concept of human capital has evolved beyond simple measures of educational attainment to encompass a broader set of capabilities. While years of schooling remain important, research increasingly highlights the role of higher-order cognitive skills (such as critical thinking and problem-solving) and non-cognitive skills (including perseverance, self-regulation, and social competence) in shaping both economic outcomes and broader life trajectories (Heckman and Kautz, 2012; Lindqvist and Vestman, 2011). A growing literature shows that several of these dimensions are malleable in childhood (Alan and Mumcu, 2024; Heckman and Mosso, 2014), raising a central policy question: how to cultivate them in resource-constrained public schools. Yet instruction in such systems is typically teacher-centric and tightly syllabus-driven (Glewwe et al., 2020) and not designed to cultivate higher-order cognitive capabilities or the non-cognitive dimensions that increasingly determine long-run outcomes. In this paper, we ask whether a guided, discovery-based pedagogy that elicits curiosity and questioning can shift what we term *multi-dimensional human capital* (MDHC)—defined as curiosity, self-confidence, self-efficacy, aspirations and perceived barriers, engagement, and scholastic achievement—in low-capacity public schools in India. These dimensions are grounded in established frameworks linking motivational beliefs, behavioral engagement, and academic performance (Fredricks et al., 2004; Lam et al., 2012), and evidence that curiosity is malleable through targeted educational interventions (Alan and Mumcu, 2024), and that aspirations and perceived barriers to education can be shifted through mentoring programs (Carlana et al., 2022).

Demand for both higher-order cognitive and non-cognitive skills is rising in modern labor markets (Deming, 2017; Waddell, 2006). While much of the evidence comes from high-income settings, the changing nature of work makes these skills increasingly salient globally (World Bank, 2018). Evidence from low and middle income contexts is growing: Nikolov et al. (2020) show that higher order cognitive skills predict earnings in South Africa. Adhvaryu et al. (2023) document substantial returns to soft-skills training in Indian manufacturing, and Danon et al. (2024) show that socio-emotional skills are correlated with labor earnings conditional on years of schooling in rural Pakistan and Cambodia, suggesting that non-cognitive skills carry meaningful labor market returns even in contexts characterized by routine production work. Recent experimental work identifies promising classroom mechanisms, namely teacher-mediated curiosity pedagogy and inquiry-based science, demonstrating that curiosity and scientific thinking are malleable and can enhance learning (Alan and Mumcu, 2024; Nourani et al., 2025). Yet two questions remain open. First, can such approaches shift the broader set of skills that constitute MDHC in low-capacity public systems? Second, can they be effective for students in upper-primary grades, where content is more technical and discovery-based teaching is more demanding?

We address these questions through a randomized controlled trial in the Indian state of Uttar Pradesh, partnering with Agastya International Foundation to layer a guided, discovery-based science module on top of business-as-usual instruction in grades 6, 7 and 8.[1] We test two delivery models: a high-intensity *Mobile Science Lab* (MSL), consisting of a van equipped with a portable science lab and three instructors visiting public schools monthly for four-hour sessions, and a lower-intensity *Lab on Bike* (LoB), consisting of one instructor with portable kits visiting monthly for two-hour sessions. Both models instantiate the same design principle of guided discovery anchored in hands-on tasks, while varying intensity and staffing—a design feature that permits us to identify how delivery intensity, rather than curriculum content, shapes human capital formation. We randomly assigned 27 public schools to MSL (28 controls) and 41 public schools to LoB (36 controls), covering nearly 14,000 students across 132 schools.[2]

Sessions are highly interactive and student-centered: instructors pose simple science problems aligned to the school science curriculum, guide hands-on experimentation, connect concepts to real-life examples, and encourage questioning throughout. In a typical public school classroom in India, the teacher controls the majority of talk time; by contrast, students lead discussion for roughly 75 percent of session time in our intervention, a sharp departure from traditional pedagogy. The science focus provides a structured canvas for inquiry and directly responds to declining engagement with science in Indian schools, where ASER Centre (2024a) documents that foundational learning gaps persist into upper-primary grades and science instruction remains heavily theory-based with limited laboratory activity (ASER Centre, 2024b; Pareek, 2019).

We pre-specified three outcome families following the contextual model of student engagement in Lam et al. (2012), a prominent contextual framework that organizes motivational, behavioral, and achievement outcomes in educational research: *personal factors* (curiosity, self-confidence, self-efficacy, aspirations, perceived barriers), *engagement* (interest in science, self-assessed performance, perceptions of science teachers, attendance, enjoyment of schoolwork), and *achievement* (test scores). We estimate intention-to-treat effects separately for MSL and LoB. We also examine whether improvements in personal factors translate into greater engagement and, in turn, higher achievement, providing insight into the mechanisms through which discovery-based pedagogy operates.

The high-intensity MSL produced substantial gains across multiple dimensions of MDHC. Personal factors improved: curiosity by 0.18 sd, self-confidence by 0.16 sd, aspirations by 0.15 sd, and perceived barriers for further education by −0.23 sd, with a suggestive increase in self-efficacy. Science-specific engagement also increased, including interest in science, perceived understanding of the science teacher, and time spent on science outside school. General engagement rose as well, with higher enjoyment of school and classroom participation across subjects, without changes in attendance. Achievement improved across all

[1]Agastya International Foundation, established in 1999, is a leading Indian NGO recognized for its innovative approach to science education. As former President of India Dr. A.P.J. Abdul Kalam (also an aerospace scientist, known for his contribution in India's satellite launch vehicles) noted: "The lesson we derive from [the Agastya] experience is that innovative and student-friendly solutions are needed to enable scientific learning in the youth, especially those in rural and remote regions of the nations of the world." Agastya's vision is "to build a new India of tinkerers, creators, innovators and solution-seekers."

[2]AEA RCT Registry: https://www.socialscienceregistry.org/trials/10902.

state-administered test subjects (Science, Mathematics, English, and Hindi)[3], though our more conceptually demanding assessment and a creativity task showed no average effects. Mediation patterns are consistent with improvements in personal factors feeding into engagement; by contrast, the science score gains persist even after controlling for these mediators, consistent with a direct learning channel during lab sessions. Spillovers to non-science subjects are consistent with both a curiosity-based learning-state amplifier operating across domains and test-taking dynamics; the design does not permit us to disentangle these interpretations.

We find *no systematic changes* in teacher pedagogy, indicating that effects operate directly through the module rather than through shifts in regular classroom instruction. Broader personality traits (Big Five) and sociability remain largely unchanged, apart from a small decline in conscientiousness, suggesting that effects concentrate on education-relevant skills and behaviors rather than personality dimensions.

By contrast, the lower-intensity LoB generated limited average effects: personal factors and engagement did not move, state-administered test score gains were generally absent (apart from a small, non-robust increase in English), and there were no detectable impacts on our conceptual assessment, creativity, or broader non-cognitive outcomes. The contrast between MSL and LoB underscores the importance of sufficient intensity for discovery-based pedagogy to shift outcomes at scale in low-capacity public schools.

Heterogeneity analyses exploiting variation in program exposure and design features point to two distinct channels. An indirect motivational pathway runs through repeated exposure: curiosity, confidence, and engagement shift gradually with cumulative sessions. A direct cognitive pathway operates more immediately: heightened attention during hands-on sessions improves encoding and content retention without requiring the same cumulative activation. Within MSL, gains in personal factors and engagement emerge only after four to five sessions (improvements of 0.15–0.25 sd in curiosity, self-confidence, aspirations, and classroom engagement) with little movement at lower exposure levels. Survey-based test performance does not display a comparable gradient, consistent with content acquisition responding more immediately to exposure than requiring cumulative activation. In LoB, effects concentrate among students completing all six sessions, who experience gains of 0.18–0.24 sd across curiosity, self-confidence, engagement, and test performance, consistent with per-session intensity shaping the cumulative exposure required for motivational responses to materialize. Program features also shape effectiveness: students valuing hands-on experiments exhibit gains of 0.20–0.30 sd in personal factors and 0.15–0.18 sd in test performance, consistent with experiments raising both informational surprise and attentional focus, while valuing instructional style predicts improvements in personal factors but not test performance, consistent with a predominantly motivational pathway. Real-life examples display null or negative associations across both arms. Because session attendance and program preferences are endogenous, these within-arm patterns are descriptive; the primary causal evidence that intensity shapes effectiveness comes from the contrast between MSL and LoB.

We also examine heterogeneity by student background and instructional context. MSL effects are broadly similar across gender, pre-intervention ability, and household wealth. For LoB, lower-asset students experience larger test gains (0.19 sd on the survey-based test score, 0.16 sd on rote, and 0.17 sd on application) suggesting

[3]These are state administered tests that are different from the usual school administered tests for student assessment.

that cognitive gains concentrate among students lacking home inputs. However, lower-ability students exhibit negative impacts on performance outcomes, which imply that shifts in motivation require sufficient baseline knowledge. Gender heterogeneity does not map cleanly onto either pathway: MSL performance gains are somewhat larger for boys (−0.16 sd interaction on the test score for girls) while LoB engagement gains are somewhat larger for girls (+0.17 sd on the engagement index), but neither pattern suggests a clear mechanism. The clearest heterogeneity emerges by teacher quality: MSL produces gains of 0.48–0.60 sd on personal factors and engagement in low-teacher-quality schools but near-zero effects where regular instruction is strong, consistent with input substitution. LoB effects are not moderated by teacher quality, consistent with lower-intensity delivery falling below the activation threshold regardless of the classroom baseline. Taken together, the consistency of patterns across exposure gradients, program preferences, student background, and instructional context provides internally coherent evidence that discovery-based pedagogy operates through at least two distinct mechanisms, though a formal causal decomposition lies beyond what our design can establish.

We further benchmark MSL's effect sizes against other rigorously tested education interventions. MSL's test score gains of 0.22–0.34 sd and non-cognitive improvements of 0.15–0.18 sd place it in the mid-to-upper range of interventions tested in comparable contexts. These effects match the upper end of remedial tutoring (Banerjee et al., 2007) and are comparable to technology-aided afterschool instruction (Muralidharan et al., 2019), while falling well below intensive teacher-transformation models (Nourani et al., 2025). MSL's science test score gain of 0.28 sd sits within the historical meta-analytic range of 0.26–0.35 sd from inquiry-based curricula (Bredderman, 1983; Shymansky et al., 1990; Weinstein et al., 1982) but below more recent estimates of 0.50–0.65 sd (Furtak et al., 2012; Schroeder et al., 2007), a gap that likely reflects implementation challenges in low-resource public systems where sustaining inquiry pedagogy at scale proves more difficult than in tightly controlled experiments. MSL's notable feature is its breadth: while most interventions target test scores alone, MSL simultaneously shifts curiosity (+0.18 sd), self-confidence (+0.16 sd), aspirations (+0.15 sd), and engagement (0.17–0.18 sd), outcomes that may generate long-run returns beyond what test scores predict (Kautz and Zanoni, 2024). At $5.56 per student annually, MSL achieves approximately 5.1–6.2 sd per $100, making it 1.7 to 2.0 times more cost-effective than remedial tutoring on academic outcomes alone. Cross-regional evidence further suggests that short-run estimates understate long-run returns: Bando et al. (2019) find that inquiry-based pedagogy effects grow from 0.14 sd (science) and 0.18 sd (mathematics) initially to 0.23 sd and 0.39 sd after four years, with the cost per 0.10 sd gain falling from approximately $18 (short run) to $8–$11 (long run, depending on subject) as four-year returns are taken into account. Whether MSL follows a similar trajectory remains an open question without comparable follow-up data.

This study contributes to several strands of literature. *First*, we contribute to the literature on experiential and child-centric learning, particularly in low-capacity public education systems. While Alan and Mumcu (2024) show that teacher-mediated experiential pedagogy can stimulate curiosity among primary students in Turkey, our intervention relies on specialized instructors from an external organization rather than classroom teachers. This design responds directly to challenges in the Indian context, where teacher absenteeism and limited training effectiveness constrain pedagogical reforms (Kingdon and Teal, 2010; Muralidharan, 2012; Muralidharan and Sundararaman, 2011). Our intervention also extends beyond curiosity to additional non-cognitive outcomes, including engagement and self-efficacy, and focuses on middle school students (grades 6–8), a group for which the existing evidence remains limited. Our approach complements recent evidence on child-centric education solutions—including Teaching at the Right Level (Banerjee et al., 2016) and technology-enabled personalization (Muralidharan et al., 2019)—by demonstrating that portable, externally delivered discovery-based methods can foster multidimensional human capital in low-resource public systems.

Second, we expand the scope of large-scale evaluations of school-based interventions beyond cognitive outcomes. Much of the existing literature on school inputs and quality improvements, including investments in teaching materials, infrastructure, or tutoring (Banerjee et al., 2005; Case and Deaton, 1999; Chin, 2005; Drèze and Kingdon, 2001; Duflo et al., 2015, 2012; Glewwe and Jacoby, 1994; Urquiola and Verhoogen, 2009), has primarily focused on test scores and years of schooling. Studies of inquiry-oriented pedagogy closer to our intervention (Bando et al., 2019; Nourani et al., 2025) also emphasize academic outcomes. We depart from this focus by examining how externally delivered science labs not only improve learning but also foster personal factors and engagement. We measure a broader set of skills, including curiosity, self-confidence, self-efficacy, and perceptions of barriers, addressing concerns in Heckman and Rubinstein (2001) and Jackson (2018) about the limitations of focusing solely on cognitive metrics. While non-cognitive skill interventions have shown long-term benefits in developed countries, they are typically delivered through after-school or mentoring programs (Kemple and Snipes, 2000; Kemple and Willner, 2008; Roder and Elliott, 2011). Evidence on delivering such interventions within low-resource public school systems remains limited. We provide an externally delivered, science-focused model that extends measurement beyond cognition while operating at system scale.

Finally, this study contributes to the literature on alleviating supply-side constraints in science education by examining how alternative delivery models can address complementarities between pedagogy, laboratory infrastructure, and teacher capacity. While many developing-country schools, including those in India, possess basic science laboratories and trained teachers, the effective utilization of these resources for experiential learning remains limited (Pareek, 2019). The binding constraint is not the absence of infrastructure but rather teachers' limited capacity to employ inquiry-based pedagogical approaches that transform laboratories into spaces of active investigation. Regular schoolteachers in India often lack training in hands-on experimental methods and interactive instruction, instead relying on traditional teacher-centered approaches that result in laboratories being underused or confined to rote demonstrations rather than student-driven exploration (Brinkmann, 2015; Dyer et al., 2004). The pedagogy models we evaluate directly target this complementarity problem by bundling three essential inputs: portable experimental equipment, instructors trained in experiential pedagogy, and a curriculum designed for inquiry-based learning (Jones and Stapleton, 2017). Rather than merely supplying physical resources, this approach integrates pedagogical expertise with existing school infrastructure, illustrating how instruction quality and resource utilization are mutually reinforcing.

2 Context, Intervention, Design, and Conceptual Framework

2.1 Context

Despite near-universal enrollment, foundational learning in the Indian state of Uttar Pradesh – one of the most populous states in India – remains weak. In 2024, over 95 percent of children aged 6–14 were enrolled, yet only 56 percent of grade 5 students could read a grade 2 text and 39 percent could perform basic division. Even in grade 8, numeracy deficits persist: roughly one-third can do division even though most can read a grade 2 text (ASER Centre, 2024b).[4] District-level human-development assessments of the state show that its education indicators lag national averages (Maurya et al., 2015).

The challenge is less about access than classroom practice: instruction is syllabus-driven, emphasizes rote methods, and rarely involves inquiry or experimentation. A classroom observation exercise conducted by ASER across government primary schools in eight Indian states, including Uttar Pradesh, found wall-mounted Teaching-Learning Materials (TLM) present in 71 percent of observed classrooms; yet hand-held or manipulable TLM (the kind required for hands-on science activities) was available in only 2 of 24 classrooms observed, and was absent entirely from most schools (ASER Centre, 2024a).[5]

Implementing reforms in state education, however, involves far more than merely addressing resource gaps. Kingdon and Muzammil (2009) document system-level governance issues such as the central role of teacher unions, weak local accountability, and limited performance incentives in Uttar Pradesh's public school system. Within this context, science education faces particularly acute constraints: scarce laboratory infrastructure, low teacher confidence in running experiments, and theory-heavy instruction leave students with little exposure to inquiry-based learning. These pedagogical constraints are not unique to public schools: recent evidence suggests that the apparent private-school advantages largely reflect differences in attendance and time on homework rather than superior pedagogy (Kumar and Choudhury, 2021). This makes Uttar Pradesh a natural setting to test whether structured, discovery-based science pedagogy can improve engagement and strengthen learning in under-resourced public schools.

2.2 The Intervention: Experiential Learning Pedagogy

Science instruction in rural upper-primary schools in India (grades 6–8) follows a teacher-centered, lecture-based model organized around textbook completion and rote memorization. Although official curricula include *activity boxes* designed to promote inquiry, resource constraints and weak institutional incentives leave these almost entirely unimplemented. Students thus encounter science as a body of facts rather than as a method of investigation.

The intervention was implemented in partnership with Agastya International Foundation, a non-profit organization that designs and delivers hands-on science learning experiences for students across India.

[4] ASER Centre (2024b) only measures the learning levels for rural India.

[5] TLM are instructional aids beyond the textbook (charts, posters, manipulatives, worksheets, models, science kits, and simple lab apparatus) intended to support active, student-centered learning.

The curriculum works within the existing system rather than against it: sessions supplement prescribed textbook content by transforming it into hands-on experience through three interdependent levers (Reeve, 2012; Skinner and Pitzer, 2012): *real-life relevance*, which connects abstract concepts to students' everyday environments; *curiosity stimulation*, which activates intrinsic motivation through hypothesis formation and guided experimentation; and *autonomy support*, which positions students as active constructors of knowledge rather than passive recipients. Each session follows a structured arc comprising a curiosity-provoking opener (*super start*), collaborative small-group experimentation in which students predict outcomes before observing them, and a consolidating synthesis (*super finish*), moving students through a cognitive progression from surprise and questioning to observation, reasoning, and consolidation. Topics are synchronized with the school syllabus so that the sessions act as conceptual anchors for material already taught or soon to be introduced. Simple, portable kits make the approach feasible in resource-constrained classrooms, and random supervisory visits ensure fidelity to the model. Evidence from comparable discovery-based frameworks documents gains in conceptual understanding, persistence, and intrinsic motivation (Alan et al., 2019; Alan and Kubilay, 2025; Alan and Mumcu, 2024; Bando et al., 2019).

The two delivery arms, the *Mobile Science Lab* and the *Lab on Bike*, implement an identical curriculum through the same pedagogical sequence but differ in instructor team size, instruction time, and visit frequency, generating variation in cumulative exposure without altering content (Table 1). Detailed session design is described in Appendix C.

2.2.1 Intervention Arm 1: Mobile Science Lab (MSL)

Treatment schools in this arm received the higher-intensity intervention, delivered through monthly visits by a van equipped with portable science kits and staffed by a three-member instructor team. Each instructor was primarily assigned to one grade (6, 7, or 8) and conducted two separate two-hour sessions per visit (one in the morning and one in the afternoon), following the pedagogical sequence described in Section 2.2, with flexible rotation across grades. Each student received approximately four hours of additional science instruction per month. Because sessions replaced the regularly scheduled science period (and occasionally one additional period) on visit days, the net instructional addition was slightly below the nominal four hours, increasing total science instructional time by approximately 4 percent relative to the standard 90 hours per year. The primary mechanism of change was not the additional time but the shift in learning mode, from passive listening to active experimentation. Implementation fidelity was maintained through standardized instructor training, detailed session plans with pre-specified experiments and key message charts, and unannounced monitoring visits by Agastya supervisors.[6]

2.2.2 Intervention Arm 2: Lab on Bike (LoB)

Treatment schools in this arm received a lower-intensity version of the same intervention, delivered by a single trained instructor traveling by motorbike with portable science kits. The instructor conducted one

[6]Instructors were randomly inspected by supervisors to ensure adherence to the pedagogical approach and session structure.

two-hour session per grade (6, 7, or 8) during each monthly visit. Because a single instructor could not cover all three grade sections in one day, visits typically spanned two consecutive days to ensure all students received instruction. Each student received approximately two hours of additional science instruction per month. As in MSL, sessions replaced the regularly scheduled science period and occasionally one additional period, so the net instructional addition was slightly below the nominal two hours, increasing total science instructional time by roughly 2 percent relative to the standard 90 hours per year. The LoB curriculum, pedagogical sequence (see Section 2.2), instructor training, and monitoring protocols mirrored those of MSL; the key differences were staffing intensity (one instructor versus three) and total monthly exposure time (two hours versus four). Table 1 provides a summary of both arms.

2.3 Experimental Design: Sample Selection, Randomization, and Timeline

We implement a cluster-randomized controlled trial to evaluate the impact of an experiential science pedagogy on multiple dimensions of human capital. The sampling frame consisted of schools with upper-primary (grades 6, 7, and 8) public schools in four districts of Uttar Pradesh (Ghazipur, Gonda, Lucknow, and Varanasi), selected to capture variation in socioeconomic and administrative contexts. The districts range from the peri-urban setting of Lucknow to more rural blocks in Ghazipur and Gonda, allowing the intervention to be evaluated under heterogeneous implementation conditions. The sampling frame was constructed to ensure geographic balance and operational feasibility, targeting approximately 30–40 students per instructor per session.[7]

Sample Selection and Randomization. From an initial list of 150 candidate schools, 14 were excluded prior to randomization based on pre-specified operational criteria (eight outliers: five very large schools, one very small school, and two English-medium schools; and six additional schools dropped to align with implementation capacity), yielding a pre-randomization sample of 136 schools. Randomization was conducted at the school level, stratified by geographic blocks within each district, with treatment and control schools assigned within strata. We use administrative data from India's Unified District Information System for Education Plus (UDISE+), the successor to the District Information System for Education (DISE), which has been widely used in economic research on education and infrastructure (Adukia et al., 2020; Jagnani and Khanna, 2020; Khanna, 2023), to construct baseline covariates including grade-wise enrollment, student caste composition, school infrastructure, and teacher characteristics.[8] Schools were matched to UDISE+ records using their unique 11-digit school codes. The geographic distribution of study schools is shown in Figure A.1.

Treatment modalities were assigned at the district level: selected blocks in Ghazipur and Gonda implemented MSL, while Lucknow and Varanasi implemented LoB. This assignment reflects operational constraints in deploying distinct delivery models and minimizes contamination across modalities. Within

[7]Operational thresholds guided eligibility and scale: for MSL, the minimum total upper-primary enrollment per school was 120–160 students, with a district-level cap of 2,000–2,200; for LoB, the corresponding thresholds were 60–80 students per school and 1,000–1,200 per district. These thresholds were based on prior-year enrollments to ensure logistical feasibility.

[8]UDISE+ is one of the largest education management information systems in the world, covering over 1.48 million schools, 9.5 million teachers, and 265 million students. UDISE+ is its updated digital platform, operational since 2018–19.

each district, schools were randomly assigned to treatment or control within block strata.[9] Thus, identification relies on within-district comparisons between treatment and control schools, with randomization stratified by geographic strata. The resulting distribution was: Ghazipur (30 schools: 15 treatment, 15 control), Gonda (28: 15 treatment, 13 control), Varanasi (52: 27 treatment, 25 control), and Lucknow (26: 15 treatment, 11 control). Following randomization, four schools declined participation (three MSL treatment and one LoB treatment), yielding a final analysis sample of 132 schools: 55 in the MSL arm (27 treatment, 28 control) and 77 in the LoB arm (41 treatment, 36 control). Each treatment arm is evaluated against its own control group within the same district–block strata: MSL treatment schools are compared only to MSL control schools in Ghazipur and Gonda, while LoB treatment schools are compared only to LoB control schools in Lucknow and Varanasi. The 64 control schools received no exposure to Agastya sessions, materials, or instructors; students continued standard teacher-centric, lecture-based science instruction (Section 2.2). Survey instruments, testing protocols, and data collection procedures were identical across all arms.

Timeline. The intervention was implemented from August 2022 to February 2023, with each treatment school receiving approximately one dose of their respective interventions per month over the academic year.[10] Endline data collection took place between February and March 2023, approximately two weeks after the final session.

2.4 Conceptual Framework

We organize our hypotheses around a model in which instructional context shapes learning through two distinct but interacting channels. The first is a *direct cognitive pathway*, through which hands-on experimentation improves comprehension and retention of scientific concepts by promoting active processing and conceptual integration. The second is an *indirect motivational pathway*, through which changes in students' motivational beliefs alter engagement, persistence, and effort. Following Lam et al. (2012), engagement—encompassing participation, persistence, emotional connection to learning, and cognitive investment—functions as the central mechanism translating pedagogy into student outcomes (see Figure A.2). When pedagogy activates intrinsic motivation and strengthens self-beliefs, it alters not only what students learn but how they engage with learning tasks and perceive their own competence. While both channels likely operate simultaneously, the pattern of non-cognitive shifts and component heterogeneity we document is consistent with a central role for the motivational pathway in translating experiential pedagogy into sustained learning gains.

The intervention activates the motivational pathway through three interdependent features aligned with self-determination theory (Reeve, 2012; Skinner and Pitzer, 2012): real-life relevance, which connects abstract concepts to students' everyday environments; curiosity stimulation, which activates intrinsic motivation through hypothesis formation and guided experimentation; and autonomy support, which positions students as active constructors of knowledge rather than passive recipients. Of these, curiosity stimulation is the primary mechanism through which the intervention generates informational incongruity. Curiosity is understood as a

[9] Blocks (strata) by district: Ghazipur (Ghazipur(Sadar), Manihari, Zakhania); Gonda (Jhanjhari, Mujhena); Lucknow (Chinhat, Mohanlal Ganj); Varanasi (Chiraigaon, Harahua -split into two, Sewapuri, Kashi Vidyapeeth).

[10] Implementation followed approvals from the Uttar Pradesh Basic Education Board and consent from school administrations.

drive aroused by structured incongruity—situations in which competing beliefs or expectations cannot be simultaneously satisfied—that motivates knowledge acquisition through its resolution (Berlyne, 1954). This drive arises not from novelty alone but from the perception of a specific, potentially resolvable gap between what one knows and what one wishes to know (Loewenstein, 1994); effective pedagogy must therefore make knowledge gaps salient before inquiry can take hold. Because the intervention targets these epistemic states rather than performance feedback, and because motivational orientation is not subject-specific, shifts in curiosity, self-confidence, and engagement may precede improvements in academic achievement and extend beyond science to other subjects.

Two structural features of this process inform our interpretation of the empirical patterns. The first is *temporality*: motivational and engagement shifts may emerge gradually through repeated exposure, while academic gains reflect both immediate content learning and subsequent consolidation. The second concerns the *role of instructional intensity*. Theory suggests that curiosity is unlikely to arise without a prior knowledge base; it intensifies as students accumulate partial understanding and become aware of specific gaps rather than scaling uniformly with exposure (Loewenstein, 1994). Furthermore, stimuli are most curiosity-arousing at an intermediate degree of familiarity, as too much novelty yields insufficient cognitive conflict to sustain engagement while full assimilation eliminates it (Berlyne, 1954). Together, these theoretical properties suggest that motivational responses may be non-linear in dosage, with meaningful activation of the indirect pathway potentially requiring sufficient cumulative exposure. By contrast, content learning may accumulate more steadily with attendance. We do not treat these as causal claims established by our design; session attendance is endogenous and dose-response patterns are descriptive. Rather, this theoretical backdrop motivates our focus on the contrast between MSL and LoB as the primary source of causal evidence that intensity shapes effectiveness, and informs our interpretation of the descriptive within-arm patterns as broadly consistent with these predictions.

The broader institutional context mediates how these mechanisms unfold. Because instruction is externally delivered, the intervention primarily alters students' learning experiences rather than teacher behavior or classroom pedagogy more broadly. Treatment effects may therefore vary with students' baseline ability, household resources, and participation intensity, reflecting either compensatory dynamics, where the intervention fills gaps for disadvantaged students, or reinforcing patterns, where benefits accrue to those more able to engage with experiential content. Under lower-intensity delivery, where epistemic activation depends more heavily on prior knowledge to render informational gaps salient, reinforcing patterns along baseline ability are more likely to emerge.

These two channels generate distinct empirical predictions for our three pre-specified outcome families. *Personal factors* (curiosity, self-confidence, aspirations, self-efficacy, and perceived barriers) are the proximate outcomes of the motivational pathway: they capture shifts in students' motivational beliefs and epistemic states that the intervention directly targets. *Engagement* occupies an intermediate position in the Lam et al. (2012) model, sitting downstream of personal factors (and thus a product of the motivational pathway) and upstream of achievement (and thus a proximate input to performance). Engagement gains that follow personal-factor improvements are therefore consistent with the motivational pathway feeding achievement through sustained effort and participation. *Student performance*, measured on both the researcher-administered science assessments and the state-administered standardized tests, captures the combined output of both channels: direct content acquisition during sessions via the cognitive pathway, and cumulative achievement gains mediated through motivation and engagement via the motivational pathway. This mapping guides our interpretation of the results throughout.

3 Data Collection, Outcomes Measures, and Validity of the Experimental Design

Our analysis combines administrative data with original survey data collected from students and teachers. Administrative datasets are used to implement school-level randomization, assess balance, and measure academic achievement outcomes. Survey data provide complementary information on student characteristics and post-intervention outcomes that are not captured in administrative records.

3.1 Administrative Records

3.1.1 School-Level Administrative Data (UDISE+)

We use school-level administrative data from the Unified District Information System for Education (UDISE+) to construct the sampling frame and implement school-level randomization. UDISE+ data from the 2020 rounds provide comprehensive pre-intervention information on enrollment, teacher characteristics, and school infrastructure, allowing us to ensure stable measurement of school characteristics prior to program implementation. The variables used for stratification and balance checks span three broad domains: enrollment measures, including grade-wise enrollment, the female share of students, and caste composition; teacher characteristics, capturing the number of teachers, their gender composition, and qualifications; and indicators of physical infrastructure, such as the presence of a boundary wall, the number of classrooms and buildings, availability of toilets and electricity, playgrounds and medical facilities, and the distribution of textbooks and uniforms. Together, these variables allow us to verify that treatment and control schools are comparable along dimensions that plausibly shape learning environments.

3.1.2 Student Achievement Data

We measure academic achievement using student-level administrative test score data obtained from Block Education Offices. Our primary academic outcome is based on the *Nipun Assessment Test* (NAT), a statewide, externally administered standardized assessment conducted by the Government of Uttar Pradesh. Because the NAT is designed and implemented independently of the intervention, it provides an objective measure of learning outcomes that is not subject to experimenter influence. The NAT was administered in December 2022 to students in Grades 6, 7, and 8 in Science, Mathematics, Hindi, and English. The assessment consisted of multiple-choice questions recorded on Optical Mark Recognition (OMR) sheets (see Figure F.1) and was administered under standardized protocols monitored by the state government.[11] Completed OMR sheets were digitized shortly after test completion using the official government application, ensuring timely processing and minimizing scope for manipulation.

3.2 Student Surveys

3.2.1 Pre-Intervention Student Survey

A student-level survey was administered after school-level randomization and prior to the start of the intervention (July 2022, N = 12,318) in 132 schools using OMR sheets. This survey collected information on students' demographic and socioeconomic characteristics, including gender, parental education, household assets, and household size. While comprehensive in capturing student characteristics, the survey deliberately excluded measures of primary learning or non-cognitive outcomes to minimize respondent burden and avoid priming effects. These data serve two distinct purposes: documenting the composition of the study sample and enabling robustness checks that condition on student-level characteristics.

Our primary specifications rely on randomization alone for identification, consistent with McKenzie (2012) and recent guidance from Kerwin et al. (2024) who point out that including pre-intervention controls is not required for consistency and may reduce statistical power when pre-intervention data are incomplete. Moreover, conventional balance tests at the individual level are prone to over-rejection in large samples, potentially leading to unnecessary concerns about randomization quality. We therefore present our main results controlling only for the school-level characteristics used in the randomization procedure. Specifications that additionally include pre-intervention student-level controls are reported as robustness checks.

3.2.2 Endline Student Survey

The endline student survey, administered after the intervention concluded (February–March 2023) across 131 schools ($N \approx$ 14,200), forms the primary dataset for all empirical analyses. One school from the Lab on Bike control group declined to participate in the endline data collection. The survey was conducted in classrooms by trained enumerators using pencil-and-paper instruments recorded on OMR sheets (see Appendix E.2 for a description of the variables and the scales used). The timing of this survey was chosen to coincide with the end of the academic year while minimizing disruption to regular instruction, and it provides the primary source of information on post-intervention student outcomes that are not available in administrative records.

The choice of OMR sheets was motivated primarily by cost considerations relative to the most natural alternative–tablet–or smartphone-based data collection using platforms such as SurveyCTO. Implementing digital surveys at the scale of this study would have been substantially more costly. By contrast, OMR-based

[11]The state planned to ensure a fair and transparent assessment through flying squads at the block level, strict monitoring, and cross-vigilance (Source: https://timesofindia.indiatimes.com/education/news/nat-2023-nipun-assessment-test-begins-in-council-schools-across-uttar-pradesh/articleshow/103565887.cms). The marks were scanned through the "Saral App" (mobile application) after the assessment. The total exam duration was 1:30 hours, and all subject questions were combined.

data collection allowed the per-student survey cost to be kept below $2, making it feasible to administer a detailed student survey across a large number of schools. In addition to lowering costs, the use of OMR sheets reduced data entry errors by eliminating manual transcription and enabled rapid processing of responses through automated scanning. OMR sheets also incorporated anonymous barcode identifiers that facilitated reliable linkage of student responses across survey rounds while preserving confidentiality. Taken together, these considerations made OMR-based surveys a practical and reliable approach for collecting high-quality student-level data in this context, despite the constraint that survey questions needed to be structured in a multiple-choice format.

To ensure high coverage, enumerators coordinated with schools in advance, tracked student rosters, and conducted follow-up visits in cases of above-average absenteeism.

3.3 Teacher Survey

We administered short surveys to science teachers at both the pre- and post-intervention school visits. The teacher survey collected information on demographic characteristics, educational background, teaching experience, instructional practices, and pedagogical beliefs (see Appendix E.3 for a description of the variables and the scales used). These data are used primarily to document existing instructional environments and to assess whether the intervention affected reported teaching practices. Given that we do not have full coverage of all the relevant science teachers for all the schools, we present this as suggestive evidence rather than treatment effects.

3.4 Outcome Variables

Our outcome measures follow the contextual model of engagement (Lam et al., 2012), which conceptualizes student outcomes as arising from the interaction of motivational beliefs, engagement behaviors, and academic performance. This framework allows us to examine whether experiential science instruction affects not only what students learn, but also how they perceive learning and how they engage with school. Accordingly, our primary outcomes correspond to the three families in the contextual model of engagement, while secondary outcomes capture higher-order skills and exploratory non-cognitive traits. Appendix E provides detailed survey instruments, response scales, and index construction.

3.4.1 Primary Outcomes

Primary outcomes capture changes in motivational beliefs, engagement, and science achievement directly targeted by experiential pedagogy. We organize them into three families: personal factors, engagement factors, and student performance.

Personal Factors. We measure five outcomes reported in Table 2: curiosity, self-confidence, barriers to education, aspiration, and self-efficacy. Curiosity and self-confidence are drawn from two sub-components of the Attitudes Toward Critical Thinking Scale (ATCT) developed for adolescents by Manassero-Mas et al.

(2022), which adapts elements of the California Critical Thinking Dispositions Inventory originally designed for adults. Curiosity is measured using five items on a seven-point Likert scale capturing students' intrinsic inclination to explore new ideas and enjoy the learning process (e.g., "I am curious because I try to find out what I don't know"; "I enjoy learning new things"; one item is reverse-coded). Self-confidence is measured using six items on the same scale, reflecting trust in one's own reasoning and problem-solving abilities (e.g., "I am able to understand things by myself"; "I notice that I reason better and better"; one item reverse-coded). Aspiration captures the highest level of education students intend to complete (up to grade 8, grade 10, grade 12, or a college degree), following Carlana et al. (2022). Self-efficacy measures students' perceived likelihood of achieving a college degree, independent of their stated aspiration. Barriers to education aggregates students' assessments of four potential obstacles to achieving their educational aim: economic resources, family needs and expectations, gender or caste prejudice, and not feeling academically up to the standard.

Engagement Factors. We measure six outcomes reported in Table 3, spanning affective, behavioral, and cognitive dimensions of engagement (Fredricks et al., 2004): science interest, self-perceived performance in science, understanding of the science teacher, school enjoyment, a composite engagement index, and attendance. Science interest is constructed from three items assessing students' enjoyment of science, their perception that science explains how the world works, and their interest in conducting experiments. Understanding of the science teacher is an index of five items capturing perceived clarity of instruction—whether the teacher is easy to understand, provides clear answers to questions, explains science well, uses varied teaching methods, and re-explains topics when needed. Self-perceived performance in science aggregates five items on students' academic self-concept in science, including beliefs about doing well, learning quickly, and not finding science confusing or harder than other subjects. School enjoyment is measured by a single item on students' enjoyment of going to school. The composite engagement index is adapted from Skinner et al. (2009) and aggregates items covering effort, participation, attention, listening carefully, homework completion, help-seeking, independent study, and engagement with science outside the classroom. Attendance is measured using students' self-reported number of days absent in the past week, so that lower values correspond to higher attendance.

Student Performance. We measure science learning using two complementary approaches. First, we administer curriculum-aligned science assessments distinguishing between rote-based questions that test factual recall and application-based questions that require students to apply scientific reasoning to novel scenarios.[12] Second, we use student-level administrative test scores from the *Nipun Assessment Test* (NAT), a statewide standardized assessment administered by the Government of Uttar Pradesh across Grades 6–8 in December 2022, covering Science, Mathematics, Hindi, and English. Because the NAT is designed and administered independently of the research team, it provides externally validated achievement measures not subject to experimenter influence. To assess spillovers beyond science, we additionally analyze NAT scores in Mathematics, Hindi, and English.

[12]For example, a rote-based question asks students to identify the medium in which sound travels fastest (solid, liquid, gas, or vacuum). In contrast, an application-based question asks students to rank the force of friction experienced by a toy car on different surfaces (wet marble floor, dry marble floor, newspaper, and towel), requiring conceptual application to a novel scenario.

3.4.2 Secondary Outcomes

Secondary outcomes capture broader higher-order skills and non-cognitive traits, as reported in Table 5 and 6.

Creativity. We measure creativity using a divergent thinking task based on the Alternative Uses Test. Students are asked to generate alternative uses for common objects (e.g., plastic bottles, old newspapers, and rags). We construct two measures: (i) fluency, defined as the total number of valid uses generated, and (ii) originality, based on independent evaluations by coders blind to treatment status. This non-self-reported measure captures whether experiential learning translates into enhanced creative thinking.

Sociability and Big Five Personality Traits: As exploratory outcomes, we examine whether exposure to experiential, inquiry-based instruction affects students' social behavior and personality traits related to learning and exploration. Sociability is proxied by the number of friends a student report. Personality is measured using a short inventory adapted from the Ten-Item Personality Inventory (TIPI) developed by Gosling et al. (2003), providing brief measures of the Big Five traits: extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience. These outcomes are not direct targets of the intervention and are therefore interpreted cautiously, as suggestive evidence on broader non-cognitive spillovers rather than central program impacts.

3.5 Independent Variables

Participation Intensity. Participation intensity is measured by the number of sessions each student reported attending. To minimize recall bias, enumerators listed each session by topic on the blackboard at endline, helping students identify which sessions they had attended. Among treated students, reported attendance ranged from 0 to 6 sessions.[13] The average number of sessions attended is 3.6 for MSL and 3.2 for LoB These averages include students with no or inconsistent responses, who are conservatively assigned zero sessions; excluding these students, the averages are 4.0 and 3.7, respectively.

Pedagogical Features. To examine which components of the intervention drove student engagement, we exploit variation in students' self-reported enjoyment of specific session attributes. Students identified which aspects they liked most from a pre-defined list: group activity, science experiments, real-life examples, freedom to express, and instructional style. For each attribute, we construct a binary indicator equal to one if the student reported liking it, and interact these indicators with the treatment dummy.[14]

Household Wealth. We construct a household asset index to proxy for socioeconomic status. The index aggregates information on housing quality (number of rooms and roof material), ownership of durable assets (motor vehicles, television, and phone type), and livestock holdings. These items are reported by students

[13] The full distribution for MSL is as follows: 0 sessions (58 or 2%), 1 session (384 or 13.4%), 2 sessions (315 or 11%), 3 sessions (278 or 9.8%), 4 sessions (316 or 11%), 5 sessions (470 or 16.4%), and 6 sessions (828 or 28.9%); an additional 214 students (7.5%) gave no or inconsistent responses. For LoB: 0 sessions (121 or 3.1%), 1 session (682 or 17.2%), 2 sessions (558 or 14.1%), 3 sessions (671 or 17.0%), 4 sessions (425 or 10.7%), 5 sessions (455 or 11.5%), and 6 sessions (1,045 or 26.4%); 435 students (11.0%) provided no or inconsistent responses.

[14] Students could select multiple options or indicate they enjoyed none. Among MSL students, science experiments were the most commonly liked feature (49 percent), followed by instructional style (19 percent) and freedom to express (16 percent). The pattern is similar for LoB students: science experiments again ranked first (41 percent), with instructional style (19 percent) and group activities (18 percent) cited next. Approximately 5 percent of students in both arms reported not enjoying any aspect of the intervention.

and are standard proxies for household wealth in settings where income data are unreliable or unavailable. We standardize the index relative to the control-group mean and standard deviation, and define a low-asset dummy equal to one for students below the median, which is used in heterogeneity analyses.

3.5.1 Index Construction and Social Desirability

For outcomes measured using multiple survey items, we construct standardized indices by aggregating the relevant component items and normalizing them relative to the control-group mean and standard deviation. Prior to aggregation, items are recoded as necessary to ensure consistent directionality across questions, so that higher values uniformly correspond to more favorable outcomes. Unless otherwise noted, treatment effects are reported in standard deviation units, facilitating comparison across outcomes measured on different scales and constructed from different underlying items.

Because several outcomes rely on self-reported measures, we additionally administer a short social desirability scale originally developed by Crowne and Marlowe (1960) and subsequently adapted into a reliable and valid short form by Hays et al. (1989). This measure allows us to assess whether treatment effects are driven by differential response tendencies rather than substantive changes in students' behaviors or beliefs. Social desirability scores are included in robustness analyses, and the estimated treatment effects are not sensitive to their inclusion. Appendix E.4 provides detailed item-level descriptions, response scales, and index construction for the social desirability scale.

3.6 Validity of the Experimental Design

3.6.1 Baseline Balance

We assess randomization validity by examining pre-intervention balance in school-level characteristics drawn from the 2020 round of the Unified District Information System for Education (UDISE+), which were used to stratify schools prior to assignment. These data capture three broad dimensions of the school environment: enrollment characteristics (total enrollment in Grades 6, 7 and 8, female share, and caste composition), teacher characteristics (number of teachers, gender composition, qualification levels, non-teaching staff, and student–teacher ratios), and physical infrastructure (boundary walls, buildings and classrooms, toilet facilities, playgrounds, electricity, medical facilities, and the distribution of free textbooks and uniforms per student).

Baseline balance statistics are reported in Appendix A for the final analysis sample, which excludes four schools—three in the MSL arm and one in the LoB arm—that declined to participate after randomization.[15] For the Mobile Science Lab (MSL) intervention, Appendix Table A.1 shows that treatment and control schools are well balanced across virtually all enrollment, teacher, and infrastructure characteristics. Differences in mean enrollment, gender composition, caste shares, teacher staffing, student–teacher ratios, classroom availability, and access to core facilities are small in magnitude and statistically insignificant. The only

[15] Appendix Tables B.7 and B.8 report balance for the full randomized samples of 30 treatment and 28 control schools (MSL) and 42 treatment and 36 control schools (LoB), prior to any attrition.

exceptions are the presence of concrete boundary walls and presence of toilets; these differences are isolated and do not extend to other infrastructure measures or translate into systematic differences in the composite infrastructure index.

For the Lab on Bike (LoB) intervention, Appendix Table A.2 presents a similarly detailed comparison. Treatment and control schools are closely balanced in terms of enrollment size, gender composition, caste composition, and most infrastructure indicators. At the same time, LoB treatment schools exhibit modestly higher baseline teacher staffing levels, lower student–teacher ratios, and a greater number of teachers with graduate degrees. These differences reflect variation in pre-existing teacher resources rather than student composition or physical infrastructure, and we account for them in the main empirical specification by controlling for the full set of school-level covariates.

Taken together, the balance results for both interventions are consistent with successful randomization and suggest that estimated treatment effects are unlikely to be driven by systematic pre-treatment differences between treatment and control schools.

3.6.2 Descriptive Statistics

Appendix Table A.3 reports student-level descriptive statistics for the post-randomization sample. Of the 14,203 students surveyed, 5,750 are in MSL schools (2,863 treatment and 2,887 control) and 8,453 are in LoB schools (4,392 treatment and 4,061 control). The table summarizes characteristics by intervention type and describes sample composition rather than assessing treatment–control balance.

Both samples exhibit similar demographic and socioeconomic profiles. Students are concentrated in lower grades: in MSL schools, 38% are in Grade 6, 35% in Grade 7, and 27% in Grade 8; in LoB schools, the corresponding shares are 39%, 32%, and 28%. Average student age is 12.7 years in MSL schools and 12.8 years in LoB schools, with near gender parity—approximately 53% of students are female in both samples. Parental education levels reflect the predominantly rural context: fathers have, on average, 8.5 years of schooling and mothers about 7.0 years across the two intervention areas. One notable difference is occupational composition: 37% of fathers in MSL schools are employed in agriculture, compared with 27% in LoB schools. Shares of self-employed and salaried fathers are similar across interventions, at roughly 20–21% and 10–12%, respectively. Average household size is approximately seven members, and the household wealth index is nearly identical across samples (2.85).

Overall, these descriptive statistics indicate that students across the two intervention areas come from broadly similar populations, the main exception being occupational composition, which likely reflects differences in local economic structure.

3.6.3 Implementation Fidelity and Program Exposure

Exposure to the intervention was tracked through multiple sources: (i) instructor attendance logs recording session completion and duration, (ii) student self-reports of participation captured in the endline survey, and (iii) classroom observations during unannounced monitoring visits conducted by Agastya supervisors. Based on Agastya report, overall compliance was high: 96% of scheduled MSL sessions and 94% of scheduled LoB sessions were completed as planned. Based on the survey responses, among students in treatment schools, 78% in MSL and 72% in LoB attended at least four of the six monthly sessions. No control school received any Agastya sessions, materials, or instructor visits, effectively eliminating concerns about cross-arm contamination. Instructor logs and monitoring records corroborate these attendance figures, indicating minimal reporting bias.

3.7 Pre-Analysis Plan

We adhere to our Pre-Analysis Plan (PAP), which is available and timestamped at RCT ID: AEARCTR-0010902. The PAP pre-specified our two treatment arms (Mobile Science Lab and Lab-on-Bike), the intent-to-treat estimation strategy with block fixed effects and school-level controls, cluster-robust standard errors at the school level, and our approach to multiple hypothesis testing using FDR-corrected q-values following Benjamini and Hochberg (1995). Our primary outcomes—curiosity, self-confidence, creativity, and commitment and engagement—and secondary outcomes, including content-based learning, aspirations, time use, and attendance, were specified in advance. The PAP also pre-specified heterogeneity analyses along socioeconomic characteristics, parental education, wealth, gender, grade, teacher motivation/quality, and program exposure—including dose-response patterns by sessions attended and heterogeneity by program components valued by students. Our empirical analysis follows these pre-specified elements closely.

We depart from the PAP in two respects. First, the PAP specified a single index for commitment and engagement; in the paper we disaggregate this into science-specific and general engagement components to provide a richer picture of the intervention's effects, consistent with the conceptual framework. Second, the PAP did not anticipate the availability of externally administered statewide test scores from the Nipun Assessment Test (NAT), which we incorporate as our primary achievement measure given its independence from the research team.[16]

4 Estimation Strategy

We estimate the intent-to-treat (ITT) effects of two experiential science pedagogy interventions – the Mobile Science Lab (MSL) and Lab on Bike (LoB) on student outcomes. Because each intervention was randomized against its own control group within distinct districts, we estimate their impacts separately and do not pool the two experiments. Let $TREAT_s^j$ denote treatment assignment, where $j \in \{\text{MSL}, \text{LoB}\}$, taking the value one if school s received intervention j and zero otherwise. Our baseline empirical specification is:

$$y_{isb} = \alpha_b^j + \beta^j TREAT_s^j + \gamma^j X_s + \epsilon_{isb}^j, \tag{1}$$

where i indexes students in school s located in geographic block b. The outcome y_{isb} denotes student-level

[16]See Banerjee et al. (2020) for a discussion of the costs and benefits of strict PAP adherence, recommending that the final paper be judged as a distinct object from the "results of the PAP."

measures of motivational beliefs, engagement with school, or academic performance. We include block fixed effects α_b^j because randomization was stratified by block within districts.

Because treatment is assigned at the school level, we cluster standard errors by school. To improve estimation precision, we additionally control for the vector of school-level covariates X_s used in the randomization procedure. These include standardized measures of school resources – non-teaching staff, toilets, graduate teachers, free textbooks, and free uniforms – each scaled by total enrollment in grades 6–8, as well as a school infrastructure index constructed as the sum of five binary indicators capturing the presence of a boundary wall, playground, electricity, medical checkup facilities, and a library. The coefficient of interest, β^j, identifies the causal effect of assignment to intervention j. Throughout the paper, we interpret these estimates as intent-to-treat effects and, for ease of exposition, refer to $TREAT_s^j$ simply as MSL or LoB when presenting results.

We examine heterogeneity in treatment effects along two dimensions using pre-intervention student and intervention characteristics. First, we explore heterogeneity with respect to treatment-specific characteristics Z_{isb}, such as participation intensity or students' reported liking of specific intervention attributes:

$$y_{isb} = \alpha_b^j + \beta^j TREAT_s^j + \delta^j \left(TREAT_s^j \times Z_{isb}\right) + \gamma^j X_s + \epsilon_{isb}^j. \tag{2}$$

The variables Z_{isb} are observed only for students exposed to the intervention and are set to zero for control students by construction. As such, the interaction coefficient δ^j captures differences in treatment effects associated with variation in participation or reported engagement with specific program features, but does not identify causal moderation effects due to potential selection into participation intensity. We therefore interpret these estimates as descriptive.

Second, we examine heterogeneity by predetermined student- and school-level characteristics I_{isb}, including gender, household wealth, pre-intervention academic ability, and teacher quality measured prior to the intervention:

$$y_{isb} = \alpha_b^j + \beta^j TREAT_s^j + \delta^j \left(TREAT_s^j \times I_{isb}\right) + \phi^j I_{isb} + \gamma^j X_s + \epsilon_{isb}^j. \tag{3}$$

Including the main effect ϕ^j ensures that δ^j captures differential treatment effects rather than level differences across subgroups. Teacher quality varies at the grade–school level, but because treatment assignment occurs at the school level, we continue to cluster standard errors by school throughout.

Finally, given the large number of outcomes examined, we address concerns about multiple hypothesis testing using two complementary approaches. We report family-wise error rate (FWER) adjusted p-values using Holm–Bonferroni corrections within outcome families, and false discovery rate (FDR) q-values following Benjamini and Hochberg (1995). These adjustments help distinguish robust treatment effects from chance findings arising from multiple comparisons.

5 Main Results

5.1 Student Engagement with Experiential Pedagogy

Participation rates were substantial in both treatment arms. Among surveyed students in treatment schools, 66% of MSL students reported attending at least half of the sessions (three or more out of six), compared to 60% of LoB students. Only 2% in MSL and 2.7% in LoB reported attending no sessions. These participation patterns are relevant for interpreting our intent-to-treat estimates: the main results capture average effects across all students in treatment schools, regardless of individual attendance.

The intervention was enthusiastically received by students. Among those who attended at least one session, 81% of students in the Mobile Science Lab (MSL) arm rated the sessions as *good* or *very good*, with only 9% reporting dissatisfaction. In the Lab on Bike (LoB) arm, 76% of participating students gave positive ratings, while 14% expressed dissatisfaction. When asked what they enjoyed most, students overwhelmingly highlighted the hands-on nature of the instruction: 49% of MSL students and 41% of LoB students cited science experiments as their favorite component, followed by the instructors' explanation style (19% in both MSL and LoB). Notably, 36–39% of students reported that learning through hands-on experimentation was entirely new to their educational experience, and 19–26% found the instructors' teaching approach refreshingly different from regular classroom instruction. Together, these responses suggest that the intervention successfully introduced engaging pedagogical elements that differ meaningfully from business-as-usual teaching, where practical experimentation remains limited despite the presence of activity-oriented material in textbooks.

5.2 Mobile Science Lab: Effects on Human Capital

We estimate specification 1 and report intent-to-treat (ITT) effects in standard deviation units. All outcomes are standardized using the control group mean and standard deviation.

Personal Factors. Table 2 reports effects on five personal dimensions: curiosity, self-confidence, perceived barriers to education, educational aspirations, and self-efficacy. The intervention produces consistent improvements across these outcomes. Against a control mean normalized to zero, curiosity increases by 0.18 sd and self-confidence by 0.16 sd. Perceived barriers, capturing economic constraints, social norms, and self-doubt, decline by 0.23 sd, representing a substantial reduction relative to the control distribution. Educational aspirations improve by 0.15 sd. Self-efficacy increases by 0.08 sd but is marginally insignificant (p-val=0.112).

These magnitudes are economically meaningful and comparable to behavioral interventions targeting socio-emotional skills in similar contexts (Alan et al., 2019; Alan and Mumcu, 2024). Hands-on experimentation activates curiosity through direct engagement with scientific phenomena, a process that may trigger dopaminergic reward pathways shown to enhance motivation and memory consolidation (Gruber et al., 2014). Repeated success in conducting experiments builds self-confidence by demonstrating competence and mastery. The student-centered approach may reduce perceived barriers by fostering a sense of agency and control over learning outcomes, consistent with self-determination theory (Ryan and Deci, 2000).

Engagement Factors. Table 3 presents treatment effects on academic engagement, measured at two levels: science-specific engagement and broader school engagement. Against a control mean normalized to zero, interest in science increases by 0.18 sd. Perceived understanding of the science teacher improves by 0.16 sd. Self-assessed performance in science rises by 0.15 sd but remains marginally insignificant. Complementing these classroom-based measures, Appendix Table A.44 shows that time spent studying science outside of school increases by 0.22 sd, indicating effects may have extended beyond formal instruction into students' independent learning activities.

Broader school engagement also improves: enjoyment of school increases by 0.18 sd and the classroom engagement index by 0.17 sd. Attendance is unchanged, consistent with survey timing near final examinations—a period when attendance declines across all schools regardless of treatment—rather than a failure to engage students. Because attendance remains unchanged while participation increases, the intervention operates on the intensive margin: it shifts how students interact with instruction, not merely time in school. In Indian public schools, where pedagogy is dominated by teacher-centered instruction and rote learning (Banerjee et al., 2017; Brinkmann, 2015), MSL disrupts this norm by legitimizing curiosity and active participation. These habits appear portable: gains in classroom engagement extend across subjects, consistent with improvements in personal factors—particularly curiosity and self-confidence—making students more favorably disposed toward learning generally.

Performance. We examine performance using three complementary measures: state-administered standardized tests, customized conceptual assessments, and a creativity task.

State-administered tests. Table 4 reports subject-specific treatment effects using administrative data from the Nipun Assessment Test, covering the full universe of students enrolled in the sampled grades. Outcomes are standardized relative to the control group. The MSL intervention generates sizable gains across all subjects: science scores increase by 0.28 sd, mathematics by 0.34 sd, English by 0.32 sd, and Hindi by 0.22 sd. These effects are economically meaningful and precisely estimated. We report estimates using the full universe of enrolled students rather than restricting to matched survey–administrative records in order to avoid potential bias arising from name-matching errors.[17]

The breadth of these gains is notable. Although the intervention was delivered through science instruction, improvements extend meaningfully to mathematics and language. One interpretation, consistent with the mechanism outlined in Section 2.4, is that curiosity acts as a domain-general amplifier of learning by increasing attention and receptivity to instruction across subjects (Alan and Mumcu, 2024; Gruber et al., 2014). At the same time, the Nipun assessments were conducted while the intervention was ongoing and are closely aligned with the school curriculum. As a result, the observed gains may also reflect enhanced reinforcement of curricular content during the intervention period, rather than broader or more durable changes in learning processes. An additional mechanical channel is that the integrated structure of the assessment could allow time reallocation across sections (for example, faster completion of science freeing

[17]The Nipun Assessment framework links test records to official enrollment data to reduce "ghost enrollment." Results are similar in magnitude and significance when restricting to matched survey–administrative observations; available upon request.

up time for other subjects). Our design does not allow us to separately identify these channels, and we note this as a limitation in interpreting the cross-subject gains.

Survey-based assessments. Table 5 examines performance on customized tests designed to capture conceptual understanding beyond standardized state assessments. Against a control mean of zero, we find no significant treatment effects on rote-based questions, application-based questions, overall conceptual scores, or on the creativity task measuring divergent thinking. These null effects stand in contrast to the gains of 0.22–0.34 sd observed on state-administered tests.

Two features of these assessments are important for interpretation. First, they were administered at endline, after the intervention had concluded, and therefore capture outcomes outside the period of direct instructional support. Second, they were intentionally designed to be more demanding than standard classroom assessments, requiring students to extend their learning to unfamiliar contexts and to engage in higher-order reasoning. The absence of detectable effects suggests that while the intervention improves performance on curriculum-aligned content during implementation, these gains do not immediately translate into broader conceptual understanding or transferable problem-solving skills.

Taken together, the pattern across assessments is consistent with a progression in which improvements in engagement and attention may enhance short-run mastery of taught material, while more abstract forms of learning require sustained exposure and reinforcement. At the same time, an alternative interpretation is that the test score gains primarily reflect transient reinforcement effects tied to the timing of assessment rather than durable human capital accumulation.

Non-cognitive Outcomes. Beyond academic performance, we examine whether the intervention affects broader non-cognitive dimensions. Table 6 reports treatment effects on sociability, proxied by the number of friends, and on the Big Five personality traits: extroversion, agreeableness, conscientiousness, emotional stability, and openness. We find no statistically significant impacts on sociability or on four of the five personality dimensions. The only exception is a small negative effect on conscientiousness, plausibly reflecting that the intervention's emphasis on questioning and experimentation was perceived as a departure from routine adherence. This interpretation should be treated cautiously, as the magnitude is small and unaccompanied by systematic changes in other traits. The broader pattern is consistent with experimental evidence that increasing schooling quantity does not shift core personality dimensions even in the long run (Barrera-Osorio et al., 2024), with descriptive evidence from low-income countries that years of schooling predict virtually no change in socio-emotional skills (Danon et al., 2024), and with the wider finding that core personality dimensions are relatively resistant to short-run interventions (Heckman and Kautz, 2012), whereas context-specific motivational beliefs and classroom behaviors are more malleable.

Summary. The MSL intervention generates consistent gains along the experiential learning pathway articulated in our conceptual framework: curiosity, confidence, aspirations, and classroom participation all improve, accompanied by substantial gains in state-administered test scores across subjects. Deeper analytical reasoning and stable personality traits remain unchanged, pointing to both the promise and the limits of standalone experiential interventions—they can meaningfully shift how students relate to learning and improve measurable achievement, but sustained structural inputs may be required to convert these gains into deeper cognitive transformation.

5.3 Lab on Bike: Effects on Human Capital

We estimate specification 1 and report ITT effects in standard-deviation units, standardized using the control-group mean and standard deviation.

Personal Factors. Table A.4 reports effects on curiosity, self-confidence, perceived barriers to education, educational aspirations, and self-efficacy. In sharp contrast to the MSL findings, LoB produces no discernible improvements across any dimension. Point estimates are small and statistically insignificant: curiosity (−0.019), self-confidence (0.024), aspirations (0.016), self-efficacy (−0.048), and barriers (0.046). The conceptual framework predicts that motivational shifts require threshold levels of cumulative exposure; below the threshold, epistemic incongruity is insufficient to sustain exploratory drive (Berlyne, 1954; Loewenstein, 1994). With half the monthly instructional time and a single instructor, LoB falls systematically short of this threshold on average.

Engagement Factors. Table A.5 presents treatment effects on engagement. Point estimates are uniformly small and statistically insignificant: science interest (0.021), self-perceived performance (−0.069), understanding of the science teacher (−0.068), school enjoyment (−0.010), and the composite engagement index (−0.046) are all near zero. This pattern is consistent with the predictions of the contextual model (Lam et al., 2012): since engagement is downstream of personal factors in the framework, the absence of shifts in curiosity, self-confidence, and aspirations documented above suggests that the motivational preconditions for sustained engagement were not activated. Appendix Table A.45 further shows that LoB reduces time spent on science and other subjects at tuition by 0.34 sd and 0.35 sd respectively, with no compensating gains in home study time, suggesting that the intervention did not shift students' independent learning behaviors outside the classroom. Attendance is likewise unchanged: once we control for days between the survey and final examinations, the raw estimate of −0.068 is small and insignificant.[18] Without the intensity and novelty of the MSL format, LoB falls below the activation threshold required for the motivational pathway to operate, leaving both personal factors and downstream engagement unchanged on average.

Performance. We examine performance using three complementary measures: state-administered standardized tests, customized conceptual assessments, and a creativity task.

State-administered tests. Table A.6 reports subject-specific ITT effects from the Nipun Assessment Test for all 10,289 enrolled students.[19] Point estimates are positive across subjects, namely total (0.21), science (0.12), mathematics (0.14), and Hindi (0.13), but none is statistically significant at conventional level. English (0.27) is marginally significant at the 10 percent level.

Survey-based assessments. Table A.7 shows null effects on the researcher-administered science test

[18] Endline surveys were conducted in control schools prior to treatment schools to ensure all Agastya sessions were completed before data collection. This protocol meant treatment schools were surveyed, on average, closer to final examinations, a period when attendance declines across all schools regardless of treatment. Controlling for days between survey date and final examination date eliminates the negative attendance effect in the LoB arm, confirming it was a survey-timing artifact rather than a treatment effect.

[19] We were able to match only 4,929 surveyed students to the administrative records; results using the matched sample are similar and available on request.

(0.032), its rote and application sub-scores (0.030 and 0.000), and the divergent-thinking creativity task (−0.136). Intent to choose the science stream is also unchanged (0.026). These results corroborate the administrative score findings: LoB did not generate meaningful learning gains at the level of average ITT effects.

Non-Cognitive Outcomes. Table A.8 reports treatment effects on sociability and the Big Five personality traits. Consistent with the null pattern elsewhere, we find no significant impacts on any dimension except agreeableness (−0.119, $p < 0.05$).[20] The remaining traits, namely extraversion, conscientiousness, emotional stability, and openness, as well as the number of friends, are unaffected.

Summary. LoB produces uniformly small, largely insignificant effects across personal factors, engagement, test scores, and non-cognitive outcomes. Both treatment arms delivered the same curriculum through the same pedagogical sequence; the differences were staffing intensity (one instructor versus three), monthly exposure (two hours versus four), and delivery salience. In low-capacity public school systems, delivering experiential pedagogy at a scale sufficient to sustain novelty and cognitive engagement is necessary, not merely beneficial, for the motivational pathway to activate.

5.4 Are there Spillovers on Regular Teachers?

A natural concern is whether the intervention affected regular classroom instruction, either by crowding out teaching effort or by generating positive spillovers through instructor demonstration effects—given that Agastya instructors modeled inquiry-based methods in the same school environment. Tables A.9 and A.10 examine this by estimating treatment effects on six dimensions of science teachers' self-reported pedagogical practices: instilling belief in students' abilities, instilling motivation, promoting critical thinking, use of science experiments, use of alternative examples, and use of digital equipment. Across all six dimensions and both arms, estimated coefficients are small and statistically insignificant. We find no evidence that MSL or LoB altered how regular science teachers instructed their classes. This null result is informative: it indicates that the intervention operated directly through students' learning experiences during Agastya sessions rather than by transforming the regular instructional environment. While the limited teacher-level sample—45 science teachers in MSL schools and 50 in LoB schools—constrains statistical power, these null results should be interpreted as ruling out large effects on teaching practice rather than effects of any magnitude. The achievement and engagement gains documented above therefore reflect student-level responses to the intervention itself, not downstream changes in teacher behavior.

5.5 Mechanisms: Evidence from Heterogeneity

The heterogeneity results provide evidence on the comparative statics of the conceptual framework. The patterns admit a parsimonious interpretation in which two distinct channels operate: an indirect motivational pathway, through which repeated exposure shifts curiosity, confidence, and engagement; and a direct cognitive

[20] We again obtain a negative effect on agreeableness. As discussed above for MSL, the intervention's emphasis on questioning and peer debate may lead students to perceive themselves as less deferential, reducing self-reported agreeableness.

pathway, through which heightened attention during sessions improves encoding and short-run content retention. Because session attendance is endogenous, the within-arm dose–response patterns discussed below are descriptive; we interpret them as broadly consistent with the framework rather than as causal estimates. The primary causal evidence that intensity shapes effectiveness comes from the contrast between MSL and LoB.

In terms of the outcome mapping laid out in the conceptual framework, personal factors are the proximate outcomes of the motivational pathway, engagement occupies an intermediate position as both a product of motivational shifts and a proximate input to performance, and student performance reflects the combined output of both pathways. The heterogeneity patterns below are interpreted against this structure.

Exposure gradients. Theory predicts that motivational responses require sufficiently repeated exposure to generate persistent informational gaps, whereas attention responds contemporaneously to instructional novelty. The dose–response patterns are consistent with this prediction. For MSL, statistically significant gains in personal factors and engagement emerge only after four to five sessions (improvements of 0.15–0.25 sd in curiosity, self-confidence, aspirations, and classroom engagement) with little movement at lower exposure levels (Tables 7 and 8). Survey-based test performance does not display a comparable gradient (Table 9), consistent with content acquisition responding more immediately to exposure than requiring cumulative activation.[21]

The LoB patterns sharpen this interpretation. Rather than a smooth gradient, LoB exhibits a threshold pattern: effects at one to four sessions are null or negative, while students completing all six sessions experience gains of 0.18–0.24 sd across curiosity, self-confidence, engagement, and test performance (Tables A.11, A.12, and A.13). The contrast between arms reinforces this reading: higher-intensity delivery generates broad average effects, while lower-intensity delivery produces gains only at the highest cumulative exposure.

Pedagogical features. The framework predicts that components increasing informational incongruity should affect both attention and motivation, whereas components structuring classroom interaction should operate mainly through motivation. Student-reported salience aligns with this distinction. Among MSL students, those valuing hands-on experiments exhibit gains of 0.20–0.30 sd in personal factors and 0.15–0.18 sd in test performance (Tables A.14, A.15, and A.16), consistent with experiments raising both informational surprise and attentional focus. Valuing instructional style predicts comparable improvements in personal factors and engagement but not in test performance (Tables A.14, A.15, and A.16), consistent with a predominantly motivational pathway. In the LoB arm, instructional style is the dominant positive predictor across personal factors and engagement, with gains of 0.29–0.36 sd, while its association with test performance is mixed: null on the overall score but marginally significant on the application component (Tables A.26, A.27, and A.28). Science experiments predict gains in both engagement and test performance in the LoB arm, consistent with the dual-pathway pattern observed in MSL. Real-life examples display null or negative associations across both arms and all outcome domains (Tables A.14, A.16, A.26, and A.28), suggesting that

[21]We do not report heterogeneity results for the state administered results as the total number of sessions attended was captured after the end of intervention, in the endline survey conducted in Feb/March 2023, while the state administered exam happened in December 2022.

the real-life relevance lever increases contextual familiarity without generating the informational incongruity needed to activate either pathway. Because component salience is endogenous, these patterns should be interpreted as consistent with pathway differentiation rather than as causal evidence.

Background heterogeneity. The framework further implies that motivational responses depend on sufficient prior knowledge to render informational gaps salient, whereas attentional responses may be strongest among students lacking alternative instructional inputs. For MSL, neither wealth nor pre-intervention achievement moderates effects (Tables A.17, A.18, A.19, A.23, A.24, and A.25), consistent with high-intensity delivery substituting for household learning complements and cognitive scaffolding. For LoB, lower-asset students experience larger test gains (0.19 sd on the survey-based test score, 0.16 sd on rote, and 0.17 sd on application) while engagement and personal factor interactions are null (Tables A.29, A.30, and A.31). Lower-ability students in the LoB arm exhibit negative interactions on performance outcomes (−0.13 sd on rote, −0.18 sd on creativity) with null interactions on personal factors and engagement (Tables A.35, A.36, and A.37), whereas MSL ability interactions are uniformly null (Tables A.23, A.24, and A.25). Under lower-intensity delivery, cognitive gains concentrate among students lacking home inputs, while motivational shifts require a prior knowledge base sufficient to render specific informational gaps salient. Gender heterogeneity does not map cleanly onto either pathway (Tables A.20, A.21, A.22, A.32, A.33, and A.34): MSL performance gains are somewhat larger for boys while LoB engagement gains are somewhat larger for girls, but neither pattern admits a clear mechanism interpretation.

Instructional context: teacher quality. The clearest heterogeneity in the analysis emerges by school-level teacher quality. For MSL, treatment effects on personal factors are large and significant in low-teacher-quality schools (0.48 sd on curiosity, 0.60 sd on self-confidence, and −0.45 sd on perceived barriers) but near zero or negative in high-teacher-quality schools (Table A.38). The engagement index interaction reaches 0.60 sd in low-teacher-quality schools, with additional significant interactions on perceived performance in science, teacher understanding, and school enjoyment (Table A.39). Test performance interactions are positive but imprecisely estimated (Table A.40), consistent with the motivational pathway dominating in this context. This pattern points to input substitution: where regular science instruction is weaker, MSL fills a larger instructional gap, activating motivational responses that competent classroom teachers already supply. These results complement the null effects on teacher pedagogy in Tables A.9 and A.10: the intervention operated directly through students' learning experiences rather than by changing regular instruction, and its motivational impact was largest where that baseline was weakest. For LoB, teacher quality interactions are small and statistically insignificant across all outcome domains (Tables A.41, A.42, and A.43), consistent with lower-intensity delivery falling below the activation threshold regardless of the classroom baseline.[22]

Synthesis. Across all four dimensions of heterogeneity, the patterns are internally coherent: high-intensity delivery activates both pathways broadly, low-intensity delivery activates neither on average, and MSL effects are largest precisely where the motivational baseline is weakest. These findings are consistent with the mediation interpretation in the conceptual framework, though a formal causal decomposition lies beyond

[22]Parental education does not meaningfully moderate effects in either arm across any outcome domain, consistent with teacher quality rather than household background driving the instructional context result. These results are available on request.

what our design can establish.

5.6 Robustness of the Main Results

We subject our main findings to two sets of robustness checks: controlling for social desirability bias in self-reported outcomes, and controlling for household and student characteristics.

Social Desirability Bias and Content-Specific Priming. A first concern is whether treatment effects on self-reported outcomes reflect genuine attitude change or differential response tendencies. We re-estimate all specifications controlling for an individual-level social desirability bias (SDB) index constructed using the short-form Marlowe–Crowne Social Desirability Scale (see Appendix Section E.4). Appendix Table B.1 reports results for MSL effects on personal factors, Table B.2 for engagement outcomes, and Table B.3 for non-cognitive measures; the corresponding LoB results are in Tables B.4, B.5, and B.6. The SDB index enters significantly across most self-reported outcomes and negatively for perceived barriers, as expected. Crucially, MSL treatment coefficients on personal factors, engagement, and non-cognitive outcomes are essentially unchanged in magnitude and statistical significance, indicating that the estimated effects are not driven by general response tendencies. LoB treatment effects remain small and insignificant after controlling for SDB.

A related and more specific concern is content-specific priming: because curiosity and questioning were central to the intervention, treated students may have learned that these are desirable traits and reported accordingly. Several patterns suggest that priming alone is unlikely to account for the results. First, LoB students were exposed to the same curriculum and messaging, yet show uniformly null effects on personal factors. If reporting responses were driven primarily by priming, similar directional effects should appear in LoB, albeit potentially attenuated given its lower intensity. Second, the pattern of effects is inconsistent with a generalized reporting tendency: priming would predict upward pressure across all constructs emphasized during sessions, whereas the estimated effects are concentrated in education-relevant motivational factors and do not extend uniformly across outcomes. Third, recent evidence on experimenter demand finds that even strong and differentially induced demand is too small in magnitude to generate false positives for directional effects in typical samples (Winichakul et al., 2025); it can inflate significance only when the true effect is a precise null and samples are very large. Our setting—a between-subject cluster-RCT in which control-village students have no exposure to the intervention content—further limits the scope for participants to infer the study hypothesis and adjust reports accordingly.

Controlling for Household and Student Characteristics. Appendix Figures B.1 and B.2 present coefficient estimates after additionally controlling for student and household characteristics — gender, age, father's and mother's years of education, household size, household wealth index, and social desirability bias index — alongside block fixed effects and school-level controls for pre-intervention characteristics. All specifications continue to cluster standard errors at the school level. The MSL results remain robust: positive and significant effects on personal factors (curiosity, self-confidence), engagement, and state-administered test scores (aggregate, science, mathematics, English, and Hindi) are unchanged in magnitude and statistical significance relative to the baseline specification. The LoB estimates similarly remain small and statistically indistinguishable from zero across personal factors, engagement outcomes, and test scores. These figures confirm that our main findings are not driven by imbalances in household or student characteristics.

Cheating in School Exams. Given concerns about cheating during school exams, one may worry that our school-test results are an artifact of differential cheating across treatment and control groups. However, there is no strong a priori reason to expect such differential behavior. We test this empirically following Angrist et al. (2017) and Singh (2024): we first identify clusters of classrooms with potential score manipulation (details in Appendix D), and then test whether the incidence of suspicious clusters differs between treatment and control groups. Following this approach, 10% of classes in the MSL arm and 24% of classes in the LoB arm fall into suspicious clusters, which is lower than the levels documented by Singh (2024) in Andhra Pradesh. This difference may reflect the different context, the reportedly high level of monitoring in our setting, or the absence of question-level responses in our data, which precludes the use of non-response rate variation as an additional diagnostic. Crucially, we find no statistically significant difference in the likelihood of a suspicious class being in the treatment versus the control group. As shown in Table B.9, the coefficient estimates on the suspicious class indicator are insignificant for both MSL and LoB, though positive. As a further robustness check, we re-estimate the main achievement effects after dropping suspicious classes (Tables B.10 and B.11). The gains observed in the MSL arm are robust to this exclusion. In the LoB arm, as before, there is no evidence of improvement in test scores.

Missing Outcomes. Tables B.12 and B.13 report the incidence of missing outcome variables in the MSL and LoB samples, respectively. Missing values arise from two sources: item non-response and inaccurate completion of OMR bubbles, both of which are inherent challenges in large-scale bubble-based data collection. In the MSL sample, missingness ranges from 7–12% depending on the outcome; in the LoB sample, it ranges from 9–16%. Importantly, missingness is orthogonal to treatment assignment in both arms, indicating that attrition in outcome data is not differential across treatment and control groups. Our results are therefore robust to the presence of missing values.

6 Benchmarking Effect Sizes

How does MSL compare to rigorously tested alternatives? Table 10 situates our estimates within the broader literature on education interventions in developing countries, drawing on both cognitive and non-cognitive outcomes where available.

Academic gains in context. MSL's test score gains of 0.22–0.34 sd place it in the mid-to-upper range of school-based interventions in comparable settings. The gains match the upper end of remedial tutoring (Banerjee et al., 2007) and are comparable to technology-aided afterschool instruction (Muralidharan et al., 2019), while exceeding activity-based mathematics instruction on both effect size and cost (de Barros et al., 2024). The most demanding benchmark is Nourani et al. (2025), whose teacher-transformation model in Uganda yields 0.49–0.83 sd, though through deep restructuring of regular classroom teaching rather than supplementary student-facing delivery. Within the inquiry-based science literature, MSL's 0.28 sd science gain sits within the historical meta-analytic range of 0.26–0.35 sd (Bredderman, 1983; Shymansky et al., 1990; Weinstein et al., 1982) but below more recent estimates of 0.50–0.65 sd (Furtak et al., 2012; Schroeder et al., 2007), a gap that likely reflects implementation challenges in low-resource public systems relative to tightly controlled experiments. Cross-regional evidence from Bando et al. (2019) suggests short-run gains understate long-run returns: effects of 0.14–0.18 sd compound to 0.23–0.39 sd after four years, and MSL's short-run gains of 0.28–0.34 sd already exceed those early benchmarks.

The breadth advantage. The more notable feature of MSL is not the magnitude of its academic gains but their breadth. Most interventions in Table 10 target test scores alone. MSL simultaneously shifts curiosity (+0.18 sd), self-confidence (+0.16 sd), educational aspirations (+0.15 sd), and perceived barriers (−0.23 sd), alongside engagement improvements of 0.17–0.18 sd across affective and behavioral dimensions. The nearest comparator on non-cognitive outcomes is Alan and Mumcu (2024), whose teacher-delivered curiosity curriculum in Turkish primary schools yields 0.08 sd in science and 0.11 sd in curiosity at $3.47 per student, substantially cheaper but with effects concentrated on curiosity and knowledge retention, without the broader shifts in aspirations, engagement, and achievement across subjects that MSL produces. Breadth matters for the value calculation: Kautz and Zanoni (2024) find that the OneGoal program increased college enrollment by 10–20 percentage points and reduced male arrests by five percentage points, with 13–32% of these effects mediated through non-cognitive skills. To the extent that MSL's improvements in curiosity, aspirations, and engagement similarly translate into durable changes in academic effort and attainment, cost-effectiveness ratios based on test scores alone will understate MSL's true returns.

Cost-effectiveness. Delivery cost data from Agastya International Foundation allow us to compute per-student costs directly. The MSL model costs Rs 75 per child per school visit and the LoB model Rs 60, translating to Rs 450 and Rs 360 per student annually, approximately $5.56 and $4.44 at the prevailing exchange rate (Rs 81 per USD).[23] Using the Banerjee et al. (2007) benchmark of 3.01 sd per $100, MSL achieves approximately 5.1 sd per $100 on science and 6.2 sd per $100 on mathematics, 1.7 to 2.0 times more cost-effective than remedial tutoring on academic outcomes alone. Relative to the curiosity curriculum in Alan and Mumcu (2024), at $3.47 per student, MSL costs modestly more while generating effects roughly 3.7 times larger in science and substantially broader in scope.[24] Cross-regional evidence from Bando et al. (2019) further suggests short-run estimates understate long-run cost-effectiveness: the cost per 0.10 sd gain falls from $18/$17.90 (math/science) to $8.40/$10.90 as four-year returns are taken into account.

In sum, MSL occupies a credible position on academic effect sizes and, at $5.56 per student per year, performs favorably on cost-effectiveness relative to comparable interventions. The key open question is persistence: whether the non-cognitive gains in curiosity, aspirations, and engagement documented here translate into durable improvements in educational trajectories and labor market outcomes–returns that, if they

[23] These figures reflect direct Agastya delivery costs: instructor salaries, portable equipment, and transport. They exclude curriculum development amortization, government coordination costs, and the opportunity cost of displaced school periods, and therefore represent a lower bound on full economic cost.

[24] It is also important to point out that implementing experiential pedagogy at the upper primary level requires more expensive science kits and materials since the curriculum includes relatively more advanced science concepts. Hence, a cost comparison with Alan and Mumcu (2024), which targets students at the primary school level, may not be appropriate.

materialize, could be substantial given evidence that a one-standard-deviation increase in socio-emotional skills is associated with $16.5 higher monthly earnings in comparable low-income settings (Danon et al., 2024). Establishing this through follow-up data collection would substantially strengthen the case for large-scale adoption.

7 Conclusion

This study contributes to the longstanding debate on how best to strengthen education in systems where conventional inputs—better infrastructure, more teachers, or higher salaries—are difficult to provide and slow to take effect (Chaudhury et al., 2006; Glewwe et al., 2020; Glewwe and Muralidharan, 2016). By focusing on the design of pedagogy rather than the scale of resources, it highlights that even in under-resourced schools, children's orientation toward learning can be shifted in ways that matter for future trajectories. The results underscore the idea, central to recent work in economics and psychology, that curiosity, self-efficacy, and related non-cognitive traits are not fixed endowments but can be cultivated through deliberate experiences in the classroom (Alan and Mumcu, 2024; Gruber et al., 2014; Heckman and Mosso, 2014).

Two broader lessons emerge. First, pedagogy is not a peripheral concern: how material is presented and practiced can shape students' motivation, perseverance, and confidence as much as the content itself. The evidence here resonates with programs that empower adolescents through targeted roles and peer dynamics (Alan and Kubilay, 2025), with experimental approaches that have restructured classrooms to unlock learning at scale (Banerjee et al., 2016; Muralidharan et al., 2019), and with incentive-design interventions that address supply-side barriers to effective instruction (Duflo et al., 2012). The temptation to dilute interventions for cost or logistical reasons is strong, but the contrast between MSL and LoB cautions that doing so risks undermining effectiveness altogether: a lower-intensity version of the same curriculum, delivered by a single instructor for half the monthly exposure, generated limited average effects, suggesting that pedagogical design and delivery intensity are jointly necessary conditions for impact.

Second, the findings call for moving beyond a narrow preoccupation with test scores toward cultivating multidimensional human capital. As labor markets increasingly reward adaptability, problem-solving, and socio-emotional skills (Deming, 2017), the case for educational policy that foregrounds curiosity and engagement becomes stronger. In low-capacity environments, where the promise of infrastructure and staffing reforms often lies far in the future, discovery-based pedagogy implemented with continuity offers a credible and immediate pathway.

References

Adhvaryu, A., N. Kala, and A. Nyshadham (2023). Returns to on-the-job soft skills training. *Journal of Political Economy 131*(8), 2165–2208.

Adukia, A., S. Asher, and P. Novosad (2020). Educational investment responses to economic opportunity: Evidence from Indian road construction. *American Economic Journal: Applied Economics 12*(1), 348–376.

Alan, S., T. Boneva, and S. Ertac (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *Quarterly Journal of Economics 134*(3), 1121–1162.

Alan, S. and E. Kubilay (2025). Empowering adolescents to transform schools: Lessons from a behavioral targeting. *American Economic Review 115*(2), 365–407.

Alan, S. and I. Mumcu (2024). Nurturing childhood curiosity to enhance learning: Evidence from a randomized pedagogical intervention. *American Economic Review 114*(4), 1173–1210.

Angrist, J. D., E. Battistin, and D. Vuri (2017). In a small moment: Class size and moral hazard in the italian mezzogiorno. *American Economic Journal: Applied Economics 9*(4), 216–249.

ASER Centre (2024a). Annual status of education report (ASER) 2024. Technical report, ASER Centre, New Delhi.

ASER Centre (2024b). Annual status of education report (ASER) 2024: Uttar Pradesh (rural) state findings. Technical report, ASER Centre, New Delhi.

Bando, R., E. Näslund-Hadley, and P. Gertler (2019). Effect of inquiry and problem based pedagogy on learning: Evidence from 10 field experiments in four countries. Technical Report 26280, National Bureau of Economic Research, Cambridge, MA.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukherji, M. Shotland, and M. Walton (2016). Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India. Technical report, National Bureau of Economic Research.

Banerjee, A., E. Duflo, A. Finkelstein, L. F. Katz, B. A. Olken, and A. Sautmann (2020). In praise of moderation: Suggestions for the scope and use of pre-analysis plans for RCTs in economics. Technical Report 26993, National Bureau of Economic Research.

Banerjee, A., S. Jacob, M. Kremer, J. Lanjouw, and P. Lanjouw (2005). Promoting school participation in rural Rajasthan: Results from some prospective trials. Working paper, Massachusetts Institute of Technology. mimeo.

Banerjee, A. V., 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*(4), 73–102.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics 122*(3), 1235–1264.

Barrera-Osorio, F., A. de Barros, and D. Filmer (2024). Long-term impacts of primary school scholarships: Evidence from cambodia. *Journal of Policy Analysis and Management 43*(1), 10–38.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological) 57*(1), 289–300.

Berlyne, D. E. (1954, August). A theory of human curiosity. *British Journal of Psychology 45*(3), 180–191.

Bredderman, T. (1983). Effects of activity-based elementary science on student outcomes: A quantitative synthesis. *Review of Educational Research 53*(4), 499–518.

Brinkmann, S. (2015). Learner-centred education reforms in India: The missing piece of teachers' beliefs. *Policy Futures in Education 13*(3), 342–359.

Carlana, M., E. La Ferrara, and P. Pinotti (2022). Goals and gaps: Educational careers of immigrant children. *Econometrica 90*(1), 1–29.

Case, A. and A. Deaton (1999). School inputs and educational outcomes in South Africa. *Quarterly Journal of Economics 114*(3), 1047–1084.

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*(1), 91–116.

Chin, A. (2005). Can redistributing teachers across schools raise educational attainment? Evidence from Operation Blackboard in India. *Journal of Development Economics 78*(2), 384–405.

Crowne, D. P. and D. Marlowe (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology 24*(4), 349–354.

Danon, A., J. Das, A. De Barros, and D. Filmer (2024). Cognitive and socioemotional skills in low-income countries: Measurement and associations with schooling and earnings. *Journal of Development Economics 168*, 103132.

de Barros, A., J. Fajardo-Gonzalez, P. Glewwe, and A. Sankar (2024). The limitations of activity-based instruction to improve the productivity of schooling. *The Economic Journal 134*(659), 959–984.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics 132*(4), 1593–1640.

Drèze, J. and G. G. Kingdon (2001). School participation in rural India. *Review of Development Economics 5*(1), 1–24.

Duflo, A., J. Kiessel, and A. M. Lucas (2024). Experimental evidence on four policies to increase learning at scale. *The Economic Journal 134*(661), 1985–2008.

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.

Duflo, E., R. Hanna, and S. P. Ryan (2012). Incentives work: Getting teachers to come to school. *American Economic Review 102*(4), 1241–1278.

Dyer, C., A. Choksi, V. Awasty, U. Iyer, R. Moyade, N. Nigam, N. Purohit, S. Shah, and S. Sheth (2004). Knowledge for teacher development in India: The importance of 'local knowledge' for in-service education. *International Journal of Educational Development 24*(1), 39–52.

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*(1), 59–109. Originally published 2004.

Furtak, E. M., T. Seidel, H. Iverson, and D. C. Briggs (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research 82*(3), 300–329.

Glewwe, P. and H. Jacoby (1994). Student achievement and schooling choice in low-income countries: Evidence from Ghana. *Journal of Human Resources 29*(3), 843–864.

Glewwe, P., S. Lambert, and Q. Chen (2020). Education production functions: Updated evidence from developing countries. In S. Bradley and C. Green (Eds.), *The Economics of Education* (Second ed.)., pp. 183–215. Academic Press.

Glewwe, P. and K. Muralidharan (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 5 of *Handbook of the Economics of Education*, pp. 653–743. Elsevier.

Gosling, S. D., P. J. Rentfrow, and W. B. Swann (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality 37*(6), 504–528.

Gruber, M. J., B. D. Gelman, and C. Ranganath (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron 84*(2), 486–496.

Hays, R. D., T. Hayashi, and A. L. Stewart (1989). A five-item measure of socially desirable response set. *Educational and Psychological Measurement 49*(3), 629–636.

Heckman, J. J. and T. Kautz (2012). Hard evidence on soft skills. *Labour Economics 19*(4), 451–464.

Heckman, J. J. and S. Mosso (2014). The economics of human development and social mobility. *Annual Review of Economics 6*(1), 689–733.

Heckman, J. J. and Y. Rubinstein (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review 91*(2), 145–149.

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy 126*(5), 2072–2107.

Jagnani, M. and G. Khanna (2020). The effects of elite public colleges on primary and secondary schooling markets in India. *Journal of Development Economics 146*, 102512.

Jones, A. L. and M. K. Stapleton (2017, May). 1.2 million kids and counting—Mobile science laboratories drive student interest in STEM. *PLOS Biology 15*(5), e2001692.

Kautz, T. and W. Zanoni (2024). Measurement and development of noncognitive skills in adolescence: Evidence from chicago public schools and the onegoal program. *Journal of Human Capital 18*(2), 272–304.

Kemple, J. J. and J. C. Snipes (2000, March). Career academies: Impacts on students' engagement and performance in high school. Technical report, Manpower Demonstration Research Corporation, New York, NY. ERIC Document No. ED441075.

Kemple, J. J. and C. J. Willner (2008). *Career academies: Long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood.* MDRC.

Kerwin, J., N. Rostom, and O. Sterck (2024). Striking the right balance: Why standard balance tests over-reject the null, and how to fix it. Technical Report 17217, IZA Discussion Paper.

Khanna, G. (2023). Large-scale education reform in general equilibrium: Regression discontinuity evidence from India. *Journal of Political Economy 131*(2), 549–591.

Kingdon, G. and F. Teal (2010). Teacher unions, teacher pay and student performance in India: A pupil fixed effects approach. *Journal of Development Economics 91*(2), 278–288.

Kingdon, G. G. and M. Muzammil (2009). A political economy of education in India: The case of Uttar Pradesh. *Oxford Development Studies 37*(2), 123–144.

Kumar, D. and P. K. Choudhury (2021). Do private schools really produce more learning than public schools in India? Accounting for student's school absenteeism and the time spent on homework. *International Journal of Educational Development 83*, 102395.

Lam, S.-f., B. P. H. Wong, H. Yang, and Y. Liu (2012). Understanding student engagement with a contextual model. In S. L. Christenson, A. L. Reschly, and C. Wylie (Eds.), *Handbook of Research on Student Engagement*, Chapter 19, pp. 403–419. Boston, MA: Springer.

Lindqvist, E. and R. Vestman (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics 3*(1), 101–128.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological bulletin 116*(1), 75.

Manassero-Mas, M. A., A. Moreno-Salvo, and Á. Vázquez-Alonso (2022). Development of an instrument to assess young people's attitudes toward critical thinking. *Thinking Skills and Creativity 45*, 101100.

Maurya, N. K., S. Singh, and S. Khare (2015). Human development in Uttar Pradesh: A district level analysis. *Social Science Spectrum 1*(4), 262–278.

McKenzie, D. (2012). Beyond baseline and follow-up: The case for more t in experiments. *Journal of Development Economics 99*(2), 210–221.

Muralidharan, K. (2012). Long-term effects of teacher performance pay: Experimental evidence from india. *Society for Research on Educational Effectiveness*.

Muralidharan, K., A. Singh, and A. J. Ganimian (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review 109*(4), 1426–1460.

Muralidharan, K. and V. Sundararaman (2011). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy 119*(1), 39–77.

Nikolov, P., N. Jimi, and J. Chang (2020). The importance of cognitive domains and the returns to schooling in south africa: Evidence from two labor surveys. *Labour Economics 65*, 101849.

Nourani, V., N. Ashraf, and A. Banerjee (2025). Learning to teach by learning to learn. Unpublished manuscript.

Pareek, R. B. (2019, March). An assessment of availability and utilization of laboratory facilities for teaching science at secondary level. *Science Education International 30*(1), 75–81.

Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. L. Christenson, A. L. Reschly, and C. Wylie (Eds.), *Handbook of Research on Student Engagement*, pp. 149–172. Boston, MA: Springer.

Roder, A. and M. Elliott (2011). *A promising start: Year Up's initial impacts on low-income young adults' careers*. Economic Mobility Corporation.

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*(1), 68–78.

Schroeder, C. M., T. P. Scott, H. Tolson, T.-Y. Huang, and Y.-H. Lee (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching 44*(10), 1436–1460.

Shymansky, J. A., L. V. Hedges, and G. Woodworth (1990). A reassessment of the effects of inquiry-based science curricula of the 60's on student performance. *Journal of Research in Science Teaching 27*(2), 127–144.

Singh, A. (2024). Improving administrative data at scale: Experimental evidence on digital testing in indian schools. *The Economic Journal 134*(661), 2207–2223.

Skinner, E. A., T. A. Kindermann, and C. J. Furrer (2009). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement 69*(3), 493–525.

Skinner, E. A. and J. R. Pitzer (2012). Developmental dynamics of student engagement, coping, and motivation. In S. L. Christenson, A. L. Reschly, and C. Wylie (Eds.), *Handbook of Research on Student Engagement*, pp. 21–44. Boston, MA: Springer.

Urquiola, M. and E. Verhoogen (2009). Class-size caps, sorting, and the regression-discontinuity design. *American Economic Review 99*(1), 179–215.

Waddell, G. R. (2006). Labor-market consequences of poor attitude and low self-esteem in youth. *Economic Inquiry 44*(1), 69–97.

Weinstein, T., F. D. Boulanger, and H. J. Walberg (1982). Science curriculum effects in high school: A quantitative synthesis. *Journal of Research in Science Teaching 19*(6), 511–522.

Winichakul, K. P., G. Lezama, P. Mustafi, M. Lepper, A. Wilson, D. Danz, and L. Vesterlund (2025). Effect size, experimenter demand and inference. Working paper, University of Pittsburgh.

World Bank (2018). *World development report 2019: The changing nature of work*. World Bank.

Tables

Table 1: Comparison of Mobile Science Lab and Lab on Bike Interventions

Features	**Mobile Science Lab**	**Lab on Bike**
Mode of Delivery	Van with science kits	Bike carrying science kits
Instructor Team	3 instructors + 1 driver	1 instructor
Instructor Responsibility	2 sessions per visit	3 sessions over 2 days
Instructor Quality	Same	Same
Session Time (Hrs)	2	2
Number of Sessions per student	2	1
Instruction Time (monthly)	4 hours	2 hours
Increase in Instruction Time	4%	2%

Notes: This table reports the different features of the two types of interventions - Mobile Science Lab and Lab on Bike.

Table 2: Impact of the Mobile Science Lab on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Educational Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
MSL	0.179**	0.157*	-0.231***	0.145***	0.082
	(0.077)	(0.081)	(0.068)	(0.051)	(0.051)
Family-wise error rate p-value	[0.001]	[0.000]	[0.001]	[0.000]	[0.001]
False discovery rate p-value	[0.024]	[0.059]	[0.001]	[0.006]	[0.112]
Randomization inference p-value	[0.040]	[0.074]	[0.007]	[0.016]	[0.113]
Control mean	0.000	0.000	0.000	0.000	0.000
Observations	5,750	5,750	5,750	5,386	4,994
R-squared	0.029	0.026	0.027	0.037	0.011
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports the impact of the Mobile Science Lab (MSL) treatment on student personal factors using endline survey data. The analysis sample includes 5,750 students across 55 schools (27 treatment and 28 control schools) in Uttar Pradesh. All dependent variables are normalized indices with mean zero and standard deviation one in the control group. The explanatory variable is a binary indicator for schools assigned to receive the MSL treatment. All specifications include block fixed effects (corresponding to the stratification used in randomization) and school-level controls for baseline characteristics. Column (1) reports treatment effects on the curiosity index, measured using the Hindi Epistemic Curiosity Scale. Column (2) reports effects on the self-confidence index. Column (3) reports effects on the self-perceived barriers to education index (negative coefficients indicate reduction in perceived barriers). Column (4) reports effects on educational goal aspirations. Column (5) reports effects on the self-efficacy. Robust standard errors clustered at the school level are reported in parentheses. We report three sets of adjusted p-values to account for multiple hypothesis testing: family-wise error rate (FWER) p-values in square brackets using the Romano-Wolf stepdown procedure, false discovery rate (FDR) q-values using the Benjamini-Hochberg procedure, and randomization inference p-values based on 5,000 permutations of the treatment assignment. The sample size varies across columns due to missing responses on specific survey modules. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table 3: Impact of Mobile Science Lab on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
MSL	0.182***	0.147	0.155*	0.184***	0.171**	-0.272
	(0.063)	(0.091)	(0.081)	(0.066)	(0.081)	(0.247)
Family-wise error rate p-value	[0.001]	[0.000]	[0.000]	[0.001]	[0.001]	[0.001]
False discovery rate p-value	[0.006]	[0.112]	[0.060]	[0.008]	[0.039]	[0.394]
Randomization inference p-value	[0.023]	[0.135]	[0.091]	[0.023]	[0.079]	[0.395]
Control mean	0.000	0.000	0.000	0.000	0.000	3.342
Observations	5,750	5,750	5,750	5,296	5,750	5,019
R^2	0.030	0.029	0.034	0.022	0.028	0.022
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55	55

Notes: This table reports the impact of the Mobile Science Lab (MSL) treatment on student engagement and school participation using endline survey data. The analysis sample includes 5,750 students across 55 schools (27 treatment and 28 control schools) in Uttar Pradesh. All dependent variables except attendance are normalized indices with mean zero and standard deviation one in the control group. The explanatory variable is a binary indicator for schools assigned to receive the MSL treatment. All specifications include block fixed effects (corresponding to the stratification used in randomization) and school-level controls for baseline characteristics. Column (1) reports treatment effects on the science interest index, measuring students' interest toward science subjects. Column (2) reports effects on self-perceived performance in science, capturing students' assessment of their own science abilities. Column (3) reports effects on students' understanding of their science teacher, measuring clarity of instruction and pedagogical effectiveness. Column (4) reports effects on students' enjoyment of attending school, an affective measure of school engagement. Column (5) reports effects on the composite engagement index aggregating behavioral and cognitive dimensions of participation in school and classroom activities. Column (6) reports effects on school attendance, measured as the number of days attended in the reference period (past week). Robust standard errors clustered at the school level are reported in parentheses. We report three sets of adjusted p-values to account for multiple hypothesis testing: family-wise error rate (FWER) p-values in square brackets using the Romano-Wolf stepdown procedure, false discovery rate (FDR) q-values using the Benjamini-Hochberg procedure, and randomization inference p-values based on 5,000 permutations of the treatment assignment. The sample size varies across columns due to missing responses on specific survey modules. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table 4: Impact of Mobile Science Lab on Students' Administrative Test Scores

	Aggregate	Science	Maths	English	Hindi
	(1)	(2)	(3)	(4)	(5)
MSL	0.375**	0.281**	0.342***	0.324**	0.218*
	(0.144)	(0.120)	(0.118)	(0.129)	(0.129)
Control mean	0.000	0.000	0.000	0.000	0.000
Observations	7,899	7,899	7,899	7,899	7,899
R-squared	0.205	0.179	0.150	0.153	0.172
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Grade fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports the impact of the Mobile Science Lab (MSL) treatment on students' performance in school-administered examinations using administrative test score data. The analysis sample includes 7,899 students across 55 schools (27 treatment and 28 control schools) in Uttar Pradesh. All dependent variables are normalized test scores with mean zero and standard deviation one in the control group, derived from school-conducted examinations in four core subjects. The explanatory variable is a binary indicator for schools assigned to receive the MSL treatment. All specifications include block fixed effects (corresponding to the stratification used in randomization), school-level controls for baseline characteristics, and grade fixed effects to account for differences across Grades 6–8. Column (1) reports treatment effects on the composite aggregate score index, which combines performance across all four subjects. Column (2) reports effects on the science test score, the primary subject targeted by the MSL intervention. Column (3) to Column (5) report effects on the mathematics, English and Hindi test score, respectively, examining potential spillovers to a related STEM and language subjects. The larger sample size compared to survey-based outcomes reflects the availability of administrative records for all enrolled students rather than only surveyed students. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table 5: Impact of Mobile Science Lab on Student Performance

	Science Choice	Test	Rote	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
MSL	0.011	0.004	0.058	-0.036	-0.066
	(0.034)	(0.114)	(0.106)	(0.094)	(0.091)
Control mean	0.370	0.000	0.000	0.000	0.000
Observations	5,750	5,750	5,750	5,750	5,326
R-squared	0.009	0.053	0.068	0.033	0.282
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports the impact of the Mobile Science Lab (MSL) treatment on students' test score outcomes using data from researcher-administered assessments. The analysis sample includes 5,750 students across 55 schools (27 treatment and 28 control schools) in Uttar Pradesh. All test scores and creativity scores are normalized with mean zero and standard deviation one in the control group. The explanatory variable is a binary indicator for schools assigned to receive the MSL treatment. All specifications include block fixed effects (corresponding to the stratification used in randomization), and school-level controls for baseline characteristics. Column (1) reports treatment effects on students' intent to choose science stream in further studies, measured as a binary indicator (control mean: 0.37, indicating 37% of control students intend to choose science in control sample). Column (2) reports effects on the overall science test score, aggregating performance across multiple assessment items. Column (3) reports effects on the rote learning component of the test score, capturing memorization-based knowledge. Column (4) reports effects on the application-based learning component, measuring students' ability to apply scientific concepts to novel problems. Column (5) reports effects on creativity score, assessed through open-ended tasks requiring divergent thinking and problem-solving. The smaller sample size in Column (5) reflects missing responses on the creativity assessment module. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table 6: Impact of Mobile Science Lab on Students' Non-Cognitive Measures

	Friends	Extraversion	Agreeableness	Conscientiousness	Emotional stability	Openness
	(1)	(2)	(3)	(4)	(5)	(6)
MSL	-0.128	-0.221	-0.102	-0.277**	-0.133	-0.177
	(0.148)	(0.155)	(0.128)	(0.120)	(0.118)	(0.126)
Control mean	4.350	4.360	4.490	4.650	4.740	4.470
Observations	5,145	5,166	5,167	5,169	5,168	5,169
R-squared	0.086	0.021	0.024	0.026	0.019	0.022
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55	55

Notes: This table reports the impact of the Mobile Science Lab (MSL) treatment on students' non-cognitive measures using endline survey data. The analysis sample includes 5,750 students across 55 schools (27 treatment and 28 control schools) in Uttar Pradesh. All dependent variables are measured on their original scales rather than normalized indices. The explanatory variable is a binary indicator for schools assigned to receive the MSL treatment. All specifications include block fixed effects (corresponding to the stratification used in randomization) and school-level controls for baseline characteristics. Column (1) reports treatment effects on the number of friends reported by students. Columns (2)–(6) report effects on the Big Five personality traits: Column (2) reports effects on extraversion, measuring sociability and assertiveness. Column (3) reports effects on agreeableness, capturing cooperation and compassion. Column (4) reports effects on conscientiousness, measuring organization and dependability. Column (5) reports effects on emotional stability (reverse-scored neuroticism), assessing calmness and emotional regulation. Column (6) reports effects on openness to experience, measuring curiosity and appreciation for novel ideas. The Big Five personality measures are assessed using standard personality inventory items adapted for the study context. Control means indicate the average levels in the control group, measured on the original response scales. The sample size varies across columns due to missing responses on specific personality inventory modules. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table 7: Dose–Response Effects of the Mobile Science Lab on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Educational Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
MSL × Sessions = 0	0.007	0.067	-0.149**	-0.014	-0.093
	(0.100)	(0.095)	(0.065)	(0.063)	(0.092)
MSL × Sessions = 1	0.040	-0.046	-0.214**	0.126	0.131*
	(0.120)	(0.120)	(0.105)	(0.075)	(0.074)
MSL × Sessions = 2	0.107	0.009	-0.261***	0.098	-0.043
	(0.082)	(0.092)	(0.086)	(0.087)	(0.084)
MSL × Sessions = 3	0.085	0.142	-0.174*	-0.024	-0.149
	(0.097)	(0.119)	(0.099)	(0.119)	(0.122)
MSL × Sessions = 4	0.142	0.172*	-0.237*	0.091	0.219***
	(0.120)	(0.103)	(0.127)	(0.084)	(0.079)
MSL × Sessions = 5	0.357***	0.315***	-0.368***	0.282***	0.164
	(0.106)	(0.110)	(0.113)	(0.083)	(0.110)
MSL × Sessions = 6	0.301***	0.264***	-0.218**	0.227***	0.126**
	(0.077)	(0.097)	(0.086)	(0.051)	(0.062)
Control mean	-0.000	0.000	-0.000	-0.000	0.000
Observations	5,750	5,750	5,750	5,386	4,994
R^2	0.062	0.053	0.045	0.063	0.032
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects by participation intensity in the Mobile Science Lab (MSL) intervention using endline survey data from 5,750 students across 55 schools (27 treatment and 28 control schools). The explanatory variables interact a school-level indicator for assignment to MSL with the number of sessions attended by each student (0 to 6 sessions), following specification 2. The omitted category is students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. All specifications include block fixed effects corresponding to the randomization strata, school-level baseline controls, and student-level controls. Column (1) reports effects on the curiosity index. Column (2) reports effects on the self-confidence index. Column (3) reports effects on the self-perceived education barriers index. Column (4) reports effects on students' educational aspiration to study further. Column (5) reports effects on the self-efficacy. Robust standard errors clustered at the school level are reported in parentheses. The sample size varies across columns due to missing responses on specific survey modules. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table 8: Dose–Response Effects of the Mobile Science Lab on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
MSL × Sessions = 0	-0.087	-0.009	0.033	-0.086	0.001	-0.228
	(0.084)	(0.127)	(0.116)	(0.100)	(0.094)	(0.180)
MSL × Sessions = 1	0.046	-0.066	0.024	0.074	0.005	-0.161
	(0.115)	(0.110)	(0.122)	(0.111)	(0.129)	(0.155)
MSL × Sessions = 2	0.105	0.048	-0.020	0.137	0.094	0.083
	(0.090)	(0.093)	(0.097)	(0.091)	(0.093)	(0.185)
MSL × Sessions = 3	-0.060	0.088	0.041	0.147	0.063	-0.420*
	(0.130)	(0.116)	(0.103)	(0.089)	(0.103)	(0.222)
MSL × Sessions = 4	0.058	0.170	0.110	0.081	0.183*	0.150
	(0.085)	(0.114)	(0.109)	(0.088)	(0.107)	(0.186)
MSL × Sessions = 5	0.409***	0.256**	0.352***	0.289***	0.325***	-0.014
	(0.086)	(0.111)	(0.102)	(0.098)	(0.110)	(0.146)
MSL × Sessions = 6	0.386***	0.293***	0.291***	0.322***	0.305***	-0.167
	(0.061)	(0.102)	(0.081)	(0.065)	(0.086)	(0.147)
Control mean	0.00	0.00	0.00	0.00	0.00	3.34
Observations	5,750	5,750	5,750	5,296	5,750	5,019
R^2	0.078	0.061	0.074	0.056	0.066	0.026
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects by participation intensity in the Mobile Science Lab (MSL) intervention on engagement-related outcomes using endline survey data from 5,750 students across 55 schools. The explanatory variables interact a school-level indicator for assignment to MSL with the number of sessions attended by each student (0 to 6 sessions), following specification 2. The omitted category is students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. All specifications include block fixed effects corresponding to the randomization strata, school-level baseline controls, and student-level controls.Columns (1)–(7) report treatment effects on the science interest index, self-perceived performance in science index, understanding of the science teacher index, affective engagement, composite engagement index, and self-reported attendance. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table 9: Dose–Response Effects of the Mobile Science Lab on Student Performance

	Science Choice	Test score	Rote	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
MSL × Sessions = 0	-0.088**	-0.296**	-0.157	-0.299***	-0.286***
	(0.041)	(0.121)	(0.111)	(0.102)	(0.097)
MSL × Sessions = 1	0.027	0.002	0.085	-0.086	0.043
	(0.051)	(0.140)	(0.115)	(0.132)	(0.108)
MSL × Sessions = 2	0.001	-0.019	0.039	-0.054	-0.003
	(0.046)	(0.109)	(0.105)	(0.096)	(0.121)
MSL × Sessions = 3	-0.003	-0.137	-0.028	-0.162	-0.136
	(0.053)	(0.131)	(0.133)	(0.118)	(0.092)
MSL × Sessions = 4	0.034	-0.073	0.037	-0.157	-0.138
	(0.049)	(0.131)	(0.118)	(0.135)	(0.109)
MSL × Sessions = 5	0.011	0.149	0.196	0.101	-0.078
	(0.041)	(0.126)	(0.120)	(0.114)	(0.097)
MSL × Sessions = 6	0.043	0.105	0.084	0.090	-0.020
	(0.041)	(0.146)	(0.126)	(0.125)	(0.105)
Control mean	0.370	0.000	0.000	0.000	0.000
Observations	5,750	5,750	5,750	5,750	5,326
R^2	0.020	0.078	0.086	0.050	0.296
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects by participation intensity in the Mobile Science Lab (MSL) intervention on student performance outcomes using endline survey data. The explanatory variables interact the MSL treatment indicator with the number of sessions attended (0 to 6 sessions), following specification 2. The omitted category is students in control schools. All specifications include block fixed effects corresponding to the randomization strata, school-level baseline controls, and student-level controls. Column (1) reports effects on an indicator for choosing the science stream in the future. Column (2) reports effects on the survey-based science test score. Columns (3) and (4) report effects on the rote and application-based components of the test, respectively. Column (5) reports effects on the creativity score. Robust standard errors clustered at the school level are reported in parentheses. The sample size varies across columns due to missing responses on specific modules. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table 10: Benchmarking MSL Effect Sizes Against Education Interventions

Study	Context	Intervention	Effect Size (sd)	Cost-Effectiveness
Panel A: Mobile Science Lab (Current Study)				
MSL	Uttar Pradesh, India (Grades 6–8)	Mobile science laboratories; experiential pedagogy	Math 0.34; English 0.32; Science 0.28; Hindi 0.22; Curiosity 0.18; Self-confidence 0.16; Aspirations 0.15	Rs 75/child/visit × 6 = Rs 450/yr ($5.56); **5.1–6.2 sd per $100**
LoB	Uttar Pradesh, India (Grades 6–8)	Lower-intensity mobile science; single instructor	Avg. ITT null; full-complier survey test: 0.26[†]	Rs 60/child/visit × 6 = Rs 360/yr ($4.44)
Panel B: India-Based Comparisons				
Banerjee et al. (2007)	Urban India, primary	Remedial tutoring; Computer-assisted learning	Remedial 0.14–0.28; CAL 0.35–0.47	3.01 sd per $100
Muralidharan et al. (2019)	Delhi, Grades 6–9 (4.5 months)	Technology-aided after-school instruction	Math 0.37 (ITT), 0.60 (IV); Hindi 0.23 (ITT), 0.39 (IV)	Lower cost per unit learning than govt. provision
de Barros et al. (2024)	India (13 months)	Activity-based math + teacher training	0.12 (lower-order skills)	$7.40 per student per year
Nourani et al. (2025)	Uganda	Learning to teach by learning to learn	Recall 0.49; Understanding 0.54; Reasoning 0.83; Creativity 0.79	Not reported
Panel C: International Comparisons				
Duflo et al. (2015)	Kenya, rural primary	Extra contract teacher; smaller pupil-teacher ratios	0.24 (literacy, math)	34.78 sd per $100
Duflo et al. (2024)	Ghana (500 schools; 4 interventions)	Remedial pull-out; after-school; smaller classes; tracking	Average ~0.10; High fidelity ~0.40	$126–161 initially; ~$21 at scale; 0.06–0.11 sd/$100 (0.36 at scale)
Panel D: Meta-Analyses of Inquiry-Based Science				
Furtak et al. (2012)	Meta-analysis (37 studies, K–12)	Inquiry-based science	Mean 0.50	—
Schroeder et al. (2007)	Meta-analysis (12 studies, US)	Science teaching strategies	Mean 0.65	—
Weinstein et al. (1982)	Meta-analysis (151 studies)	Science curriculum effects	Mean 0.31	—
Bredderman (1983)	Meta-analysis (57 studies)	Activity-based elementary science	Mean 0.35	—
Shymansky et al. (1990)	Meta-analysis (136 studies)	Inquiry-based curricula	Mean 0.26	—
Panel E: Cross-Regional / Non-Cognitive Interventions				
Bando et al. (2019)	Latin America (10 RCTs, 4 countries)	Inquiry and problem-based pedagogy	Short-run: Sci 0.14, Math 0.18; 4-year: Sci 0.23, Math 0.39	$18 (math) / $17.9 (science); $8.4 / $10.9 per 0.10 sd (long-run)
Alan and Mumcu (2024)	Turkey (teacher-delivered)	Curiosity-nurturing curriculum	Science 0.075; Curiosity 0.11	$3.47 per student
Kautz and Zanoni (2024)	Chicago, US (high school)	OneGoal (non-cognitive skills focus)	+10–20 pp college enrollment; –5 pp arrests (male); 15–30% mediated via non-cognitive	Not reported

Notes: Effect sizes reported in standard deviations (sd) unless otherwise noted. ITT = intent-to-treat; IV = instrumental variables accounting for attendance. Cost-effectiveness ratios report sd gains per $100 spent where available. MSL = Mobile Science Lab; LoB = Lab on Bike; pp = percentage points; Sci = Science. MSL and LoB costs are direct delivery costs from Agastya International Foundation (Rs 75 and Rs 60 per child per school visit respectively), covering instructor salaries, equipment, and transport; converted at Rs 81/USD (August 2022–February 2023). These exclude curriculum development, government coordination, and opportunity costs of displaced school time, and represent a lower bound on full economic cost. MSL cost-effectiveness of 5.1–6.2 sd per $100 corresponds to science (0.28 sd) and mathematics (0.34 sd). [†]LoB full-complier estimate is descriptive and not causal. Bando et al. (2019) long-run effects measured at four-year follow-up. Duflo et al. (2024) reports ranges across four scalable interventions. Meta-analyses (Panel D) summarize mean reported effects across studies. Nourani et al. (2025) is an unpublished manuscript.

Online Appendix

for

Science on the Move: How Experiential Pedagogy Shapes Human Capital

Nitin Kumar Bharti Samreen Malik Abhiroop Mukhopadhyay Nishith Prakash

Contents

Appendix A Appendix Figures and Tables
Appendix B Robustness Figures and Tables
Appendix C Pedagogical Design
Appendix D Survey Instruments and Measures

Table of Contents

Appendix A Appendix Figures and Tables 4

Appendix Figures

Figure A.1: Spatial Coverage of the Intervention in Uttar Pradesh 4

Figure A.2: Contextual Model of Students' Engagement 5

Balance and Descriptive Statistics

Table A.1: Pre-Treatment Balance in School Infrastructure: MSL, 2020 6

Table A.2: Pre-Treatment Balance in School-Level Characteristics: LoB, 2020 7

Table A.3: Student-Level Characteristics by Intervention Type 8

Main Results: LoB

Table A.4: Impact of the LoB on Personal Factors 9

Table A.5: Impact of the LoB on Engagement Factors 9

Table A.6: Impact of the LoB on Administrative Test Scores 10

Table A.7: Impact of the LoB on Student Performance 10

Table A.8: Impact of the LoB on Non-cognitive Measures 11

Teacher Instructional Practices

Table A.12: Did the MSL Change Teachers' Instructional Practices? 11

Table A.13: Did the LoB Change Teachers' Instructional Practices? 12

Dose–Response Effects: LoB

Table A.9: Dose-Response Effects of LoB on Personal Factors 12

Table A.10: Dose–Response Effects of the LoB on Engagement Factors 13

Table A.11: Dose–Response Effects of the LoB on Student Performance 13

Heterogeneous Effects: MSL by Program Components

Table A.14: Heterogeneous Effects by Student-Liked MSL Components on Personal Factors 14

Table A.15: Heterogeneous Effects by Student-Liked MSL Components on Engagement Factors 15

Table A.16: Heterogeneous Effects by Student-Liked MSL Components on Student Performance 15

Heterogeneous Effects: MSL by Household Wealth

Table A.17: Heterogeneous Effects of the MSL by Household Wealth on Personal Factors 16

Table A.18: Heterogeneous Effects of the MSL by Household Wealth on Engagement Factors 16

Table A.19: Heterogeneous Effects of the MSL by Household Wealth on Student Performance 17

Heterogeneous Effects: MSL by Gender

Table A.20: Heterogeneous Effects of the MSL by Gender on Personal Factors 17

Table A.21: Heterogeneous Effects of the MSL by Gender on Engagement Factors 18

Table A.22: Heterogeneous Effects of the MSL by Gender on Student Performance 18

Heterogeneous Effects: MSL by Student Ability

Table A.23: Heterogeneous Effects of the MSL by Student Ability on Personal Factors 19

Table A.24: Heterogeneous Effects of the MSL by Student Ability on Engagement Factors 19

Table A.25: Heterogeneous Effects of the MSL by Student Ability on Student Performance 20

Heterogeneous Effects: LoB by Program Components

Table A.26: Heterogeneous Effects by Student-Liked LoB Components on Personal Factors 20

Table A.27: Heterogeneous Effects by Student-Liked LoB Components on Engagement Factors 21

Table A.28: Heterogeneous Effects by Student-Liked LoB Components on Student Performance 21

Heterogeneous Effects: LoB by Household Wealth

Table A.29: Heterogeneous Effects of the LoB by Household Wealth on Personal Factors 22

Table A.30: Heterogeneous Effects of the LoB by Household Wealth on Engagement Factors 22

Table A.31: Heterogeneous Effects of the LoB by Household Wealth on Student Performance 23

Heterogeneous Effects: LoB by Gender

Table A.32: Heterogeneous Effects of the LoB by Gender on Personal Factors 23

Table A.33: Heterogeneous Effects of the LoB by Gender on Engagement Factors 24

Table A.34: Heterogeneous Effects of the LoB by Gender on Student Performance 24

Heterogeneous Effects: LoB by Student Ability

Table A.35: Heterogeneous Effects of the LoB by Student Ability on Personal Factors 25

Table A.36: Heterogeneous Effects of the LoB by Student Ability on Engagement Factors 25

Table A.37: Heterogeneous Effects of the LoB by Student Ability on Student Performance 26

Heterogeneous Effects: MSL by Teacher Quality

Table A.38: Heterogeneous Effects of the MSL by Teachers' Quality on Personal Factors 26

Table A.39: Heterogeneous Effects of the MSL by Teachers' Quality on Engagement Factors 27

Table A.40: Heterogeneous Effects of the MSL by Teachers' Quality on Student Performance 28

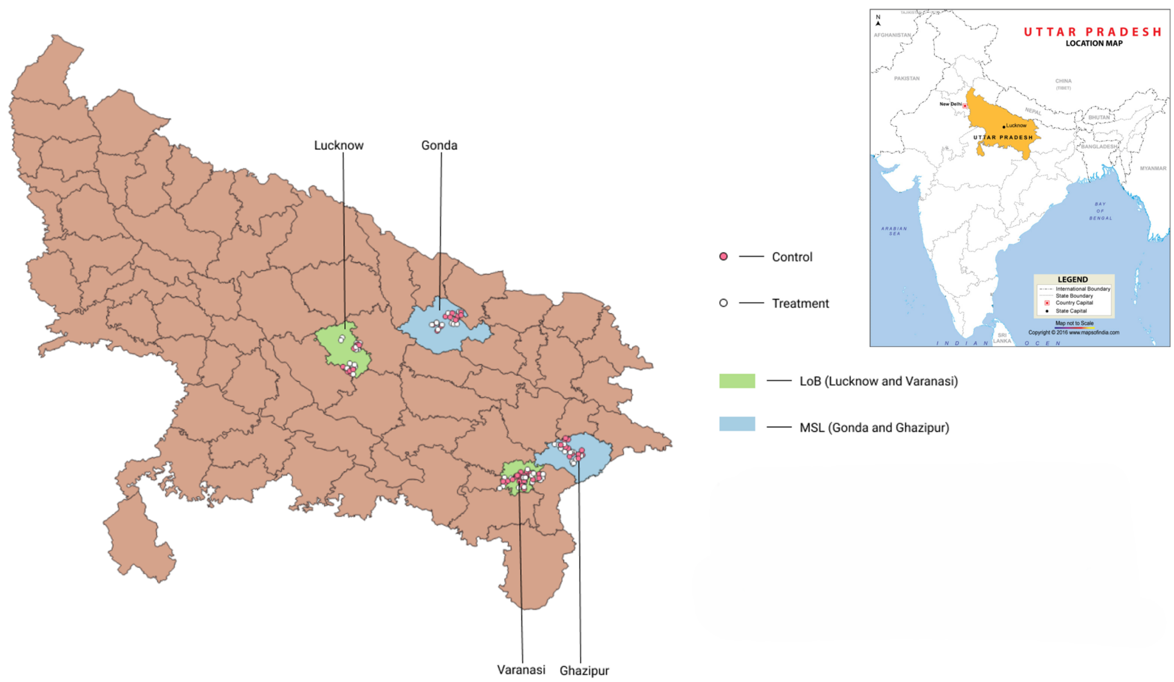
Heterogeneous Effects: LoB by Teacher Quality

Table A.41: Heterogeneous Effects of the LoB by Teachers' Quality on Personal Factors 28

Table A.42: Heterogeneous Effects of the LoB by Teachers' Quality on Engagement Factors 29
Table A.43: Heterogeneous Effects of the LoB by Teachers' Quality on Student Performance 29
Weekly Study Time
Table A.44: Impact of MSL on Weekly Study Time 30
Table A.45: Impact of the LoB on Weekly Study Time 30
Appendix B Robustness Figures and Tables 31
Robustness to Controlling Students' Characteristics
Figure B.1: Impact of the MSL: Controlling for students' characteristics 31
Figure B.2: Impact of the LoB: Controlling for students' characteristics 32
Robustness to Social Desirability Bias: MSL
Table B.1: Impact of the MSL on Personal Factors: Robustness to Social Desirability Bias 33
Table B.2: Impact of the MSL on Engagement Factors: Robustness to Social Desirability Bias 33
Table B.3: Impact of the MSL on Non-Cognitive Measures: Robustness to Social Desirability Bias 34
Robustness to Social Desirability Bias: LoB
Table B.4: Impact of the LoB on Personal Factors: Robustness to Social Desirability Bias 34
Table B.5: Impact of the LoB on Engagement Factors: Robustness to Social Desirability Bias 35
Table B.6: Impact of the LoB on Non-Cognitive Measures: Robustness to Social Desirability Bias 35
Full Pre-Treatment Balance (Original Randomized Samples)
Table B.7: Pre-Treatment Balance in School Infrastructure: MSL, 2020 36
Table B.8: Pre-Treatment Balance in School-Level Characteristics: LoB, 2020 37
Suspicious Classrooms and Administrative Test Scores
Table B.9: Are Suspicious Classes More Prevalent in Treatment Schools? 38
Table B.10: Impact of the MSL on Administrative Test Scores: Excluding Suspicious Classes 38
Table B.11: Impact of the LoB on Administrative Test Scores: Excluding Suspicious Classes 39
Table B.12: Cluster-wise Averages in MSL and LoB 44
Missing Outcomes and Treatment Indicator
Table B.12: Are missing-values correlated with the Treatment Indicator (MSL)? 40
Table B.13: Are missing-values correlated with the Treatment Indicator (LoB)? 41
Appendix C Pedagogical Design 42
Appendix D Survey Instruments and Measures 45

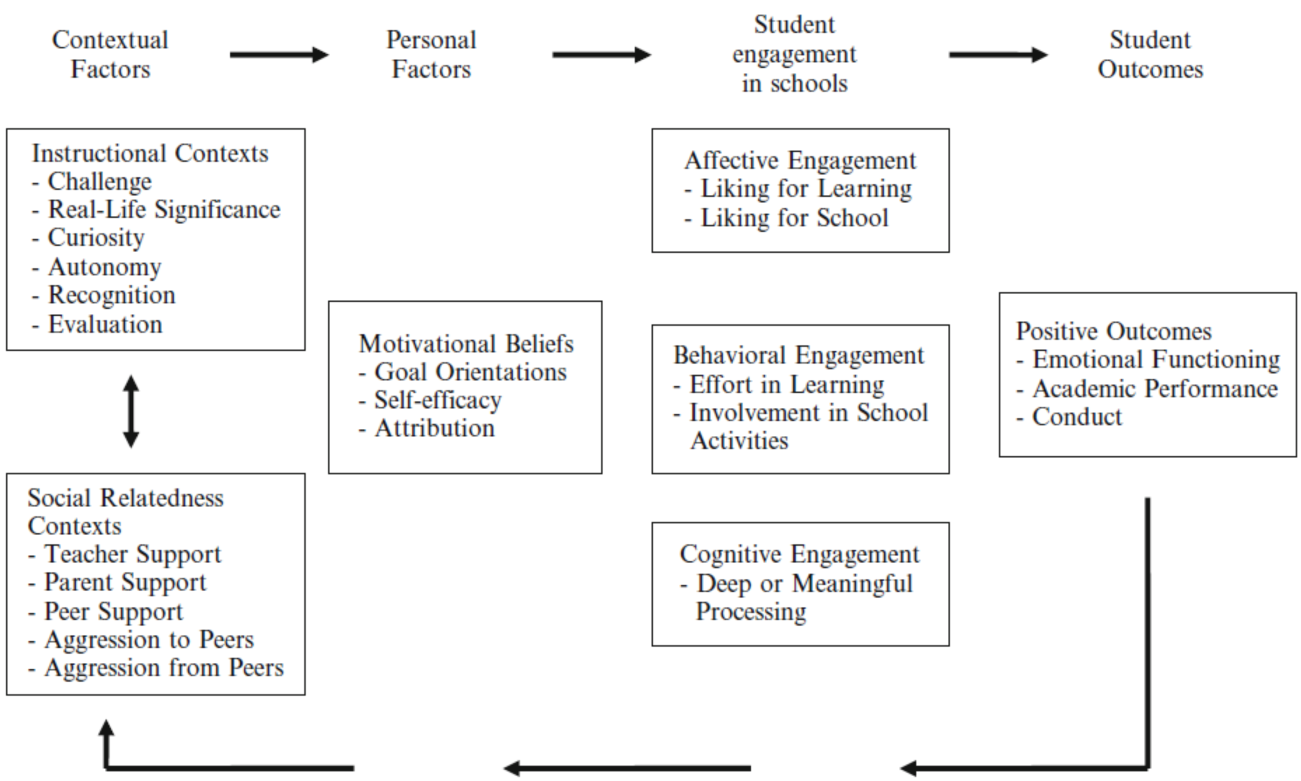
A Appendix Figures and Tables

Figure A.1: Spatial Coverage of the Intervention in Uttar Pradesh



Notes: The figure shows the spatial coverage of our intervention in the districts of Ghazipur, Gonda, Lucknow, and Varanasi in Uttar Pradesh.

Figure A.2: Contextual Model of Students' Engagement



Notes: The figure shows a simple diagram of students' model of engagement, borrowed from Lam et al. (2012), p. 406.

Table A.1: Pre-Treatment Balance in School Infrastructure: Mobile Science Lab (MSL), 2020

	Control Schools		MSL Schools		
Variable	N	Mean (SE)	N	Mean (SE)	Diff. (MSL – Control)
	(1)	(2)	(3)	(4)	(5)
Total enrollment (Grades 6–8)	28	131.68 (11.89)	27	151.74 (9.27)	20.06
Female enrollment share	28	0.51 (0.01)	27	0.49 (0.01)	-0.01
Scheduled Caste enrollment share	28	0.29 (0.03)	27	0.25 (0.03)	-0.04
Other Backward Class enrollment share	28	0.58 (0.04)	27	0.58 (0.04)	0.00
Total teachers	28	7.71 (0.76)	27	7.63 (0.76)	-0.08
Student–teacher ratio (Grades 6–8)	28	25.88 (4.60)	27	26.14 (3.31)	0.26
Female teachers share	28	0.41 (0.06)	27	0.41 (0.05)	0.00
Teachers with graduate degree or higher	28	6.89 (0.69)	27	7.22 (0.74)	0.33
Non-teaching staff	28	0.61 (0.35)	27	0.67 (0.32)	0.06
Dummy for school boundary wall	28	0.54 (0.10)	27	0.74 (0.09)	0.21
Dummy for school boundary wall (concrete)	28	0.46 (0.10)	27	0.74 (0.09)	0.28**
Number of buildings	28	2.54 (0.62)	27	2.48 (0.53)	-0.05
Number of buildings (concrete)	28	2.36 (0.60)	27	2.30 (0.53)	-0.06
Number of toilets for boys	28	1.46 (0.17)	27	1.41 (0.14)	-0.06
Number of toilets for girls	28	1.61 (0.17)	27	1.37 (0.11)	-0.24
Toilets (per student)	28	0.03 (0.00)	27	0.02 (0.00)	-0.01*
Number of classrooms	28	7.75 (0.82)	27	8.15 (0.88)	0.40
Number of classrooms (in good condition)	28	5.82 (0.75)	27	5.82 (0.77)	0.01
Dummy for school playground	28	0.75 (0.08)	27	0.78 (0.08)	0.03
Dummy for school with electricity	28	0.93 (0.05)	27	0.82 (0.08)	-0.11
Dummy for medical checkup facility	28	0.36 (0.09)	27	0.30 (0.09)	-0.06
Free textbooks distributed (per student)	28	0.61 (0.10)	27	0.57 (0.09)	-0.04
Free uniforms distributed (per student)	28	0.48 (0.09)	27	0.57 (0.09)	0.09
School Infrastructure Index	28	0.00 (0.19)	27	-0.47 (0.24)	-0.47

Notes: This table reports pre-intervention balance in school-level characteristics for the Mobile Science Lab (MSL) intervention using administrative data from UDISE+ (2020), which were used for stratification and randomization. The analysis sample includes 27 treatment and 28 control schools; three schools that declined participation after randomization are excluded. All variables are measured at the school level prior to treatment. Reported values are means with standard errors in parentheses. Column (5) reports differences in means between treatment and control schools based on two-sided *t*-tests. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.2: Pre-Treatment Balance in School-Level Characteristics: Lab on Bike (LoB), 2020

Variable	Control Schools N	Control Schools Mean (SE)	LoB Schools N	LoB Schools Mean (SE)	Diff. (LoB – Control)
	(1)	(2)	(3)	(4)	(5)
Total enrollment (Grades 6–8)	36	140.31 (9.58)	41	142.32 (7.05)	2.01
Female enrollment share	36	0.49 (0.02)	41	0.49 (0.01)	0.00
Scheduled Caste enrollment share	36	0.36 (0.03)	41	0.42 (0.03)	0.05
Other Backward Class enrollment share	36	0.57 (0.03)	41	0.54 (0.03)	-0.03
Total teachers	36	7.58 (0.79)	41	9.61 (0.71)	2.03*
Student–teacher ratio (Grades 6–8)	36	22.50 (1.73)	41	18.25 (1.61)	-4.25*
Female teachers share	36	0.64 (0.05)	41	0.64 (0.04)	0.00
Teachers with graduate degree or higher	36	6.89 (0.75)	41	8.93 (0.66)	2.04**
Non-teaching staff	36	1.17 (0.55)	41	2.39 (0.71)	1.22
Dummy for school boundary wall	36	0.83 (0.06)	41	0.83 (0.06)	0.00
Dummy for school boundary wall (concrete)	36	0.83 (0.06)	41	0.81 (0.06)	-0.03
Number of buildings	36	2.58 (0.49)	41	3.12 (0.41)	0.54
Number of buildings (concrete)	36	2.50 (0.43)	41	3.05 (0.41)	0.55
Number of toilets for boys	36	1.58 (0.14)	41	2.12 (0.22)	0.54*
Number of toilets for girls	36	1.78 (0.16)	41	2.27 (0.20)	0.49*
Toilets (per student)	35	0.03 (0.00)	41	0.03 (0.00)	0.01
Number of classrooms	36	6.81 (0.78)	41	7.66 (0.76)	0.85
Number of classrooms (in good condition)	36	5.94 (0.79)	41	6.07 (0.80)	0.13
Dummy for school playground	36	0.72 (0.08)	41	0.73 (0.07)	0.01
Dummy for school with electricity	36	0.86 (0.06)	41	0.95 (0.03)	0.09
Dummy for medical checkup facility	36	0.47 (0.08)	41	0.59 (0.08)	0.11
Free textbooks distributed (per student)	36	0.84 (0.06)	41	0.89 (0.04)	0.05
Free uniforms distributed (per student)	36	0.79 (0.07)	41	0.87 (0.05)	0.08
School Infrastructure Index	36	-0.00 (0.17)	41	0.14 (0.18)	0.14

Notes: This table reports pre-intervention balance in school-level characteristics for the Lab on Bike (LoB) intervention using administrative data from UDISE+ (2020), which were used for stratification and randomization. The analysis sample includes 41 treatment and 36 control schools; one school that declined participation after randomization is excluded. All variables are measured at baseline. Reported values are means with standard errors in parentheses. Column (5) reports differences in means between treatment and control schools based on two-sided t-tests. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.3: Student and Teacher-Level Characteristics by Intervention Type

Students	Lab on Bike (LoB)		Mobile Science Lab (MSL)	
	N	Mean (SE)	N	Mean (SE)
	(1)	(2)	(3)	(4)
Female	8,453	0.53 (0.01)	5,750	0.53 (0.01)
Grade 6	8,453	0.39 (0.01)	5,750	0.38 (0.01)
Grade 7	8,453	0.32 (0.01)	5,750	0.35 (0.01)
Grade 8	8,453	0.28 (0.01)	5,750	0.27 (0.01)
Student age (years)	8,453	12.81 (0.02)	5,750	12.69 (0.02)
Father's education (years)	8,453	8.39 (0.04)	5,750	8.64 (0.05)
Mother's education (years)	8,453	7.13 (0.04)	5,750	7.02 (0.06)
Employment (agriculture) share	8,453	0.27 (0.01)	5,750	0.37 (0.01)
Employment (self-employed) share	8,453	0.21 (0.00)	5,750	0.20 (0.01)
Employment (salaried) share	8,453	0.10 (0.00)	5,750	0.12 (0.00)
Employment (other) share	8,453	0.42 (0.01)	5,750	0.31 (0.01)
Household wealth index	8,453	2.86 (0.02)	5,750	2.85 (0.02)
Household size	8,453	6.84 (0.03)	5,750	7.19 (0.04)

Notes: This table reports student-level descriptive statistics collected post-randomization. Columns (1)–(2) correspond to students/teachers in Lab on Bike (LoB) schools, and Columns (3)–(4) correspond to students/teachers in Mobile Science Lab (MSL) schools. Reported values are means with standard errors in parentheses. Employment categories refer to the father's primary occupation. The household wealth index is a standardized asset-based measure. General Caste, Other Backward Caste and SC/ST (Scheduled Caste/Scheduled Tribe) refers to the administrative caste categories in India. No statistical tests are reported, as the table is intended to summarize baseline characteristics rather than assess treatment–control balance.

Table A.4: Impact of the Lab on Bike on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Education Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
LoB	-0.019	0.024	0.046	0.016	-0.048
	(0.057)	(0.060)	(0.037)	(0.047)	(0.047)
Control mean	0.000	0.000	0.000	0.010	-0.000
Observations	8,453	8,453	8,453	7,708	7,069
R-squared	0.013	0.018	0.020	0.011	0.005
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76

Notes: This table reports the impact of the Lab on Bike (LoB) treatment on student personal factors using endline survey data. The analysis sample includes students across 76 schools. All dependent variables are normalized indices with mean zero and standard deviation one in the control group. The explanatory variable is a binary indicator for schools assigned to receive the LoB treatment. All specifications include block fixed effects corresponding to the stratified randomization design and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels based on unadjusted t-tests.

Table A.5: Impact of the Lab on Bike on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
LoB	0.021	-0.069	-0.068	-0.010	-0.046	-0.068
	(0.061)	(0.058)	(0.057)	(0.052)	(0.056)	(0.123)
Control mean	0.000	0.000	0.000	0.000	0.000	3.309
Observations	8,453	8,453	8,453	7,433	8,453	7,331
R-squared	0.010	0.020	0.014	0.013	0.018	0.027
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76	76

Notes: This table reports the impact of the Lab on Bike (LoB) treatment on student engagement outcomes using endline survey data. The analysis sample includes students across 76 schools. All dependent variables are normalized indices with mean zero and standard deviation one in the control group. The explanatory variable is a binary indicator for schools assigned to receive the LoB treatment. All specifications include block fixed effects corresponding to the stratified randomization design and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.6: Impact of the Lab on Bike on Administrative Test Scores

	Total	Science	Maths	English	Hindi
	(1)	(2)	(3)	(4)	(5)
LoB	0.212	0.117	0.143	0.271*	0.134
	(0.128)	(0.109)	(0.116)	(0.145)	(0.107)
Control mean	0.000	0.000	0.000	0.000	0.000
Observations	10,289	10,289	10,289	10,289	10,289
R-squared	0.276	0.266	0.212	0.171	0.254
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Grade fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	72	72	72	72	72

Notes: This table reports the impact of the Lab on Bike (LoB) treatment on administrative test scores using school-conducted examinations. The analysis sample includes 10,289 students across 72 schools. All dependent variables are normalized to have mean zero and standard deviation one in the control group. The explanatory variable is a binary indicator for schools assigned to receive the LoB treatment. All specifications include block fixed effects corresponding to the stratified randomization design, school-level baseline controls, and grade fixed effects. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.7: Impact of the Lab on Bike on Student Performance

	Science choice	Test score	Rote score	Application score	Creativity
	(1)	(2)	(3)	(4)	(5)
LoB	0.026	0.032	0.030	0.000	-0.136
	(0.026)	(0.087)	(0.070)	(0.077)	(0.098)
Control mean	0.360	0.000	0.000	0.000	0.000
Observations	8,453	8,453	8,453	8,453	7,641
R-squared	0.008	0.070	0.055	0.043	0.108
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76

Notes: This table reports the impact of the Lab on Bike (LoB) treatment on student academic outcomes using endline survey data. The dependent variable in Column (1) is an indicator for intent to choose science in further studies. Columns (2)–(5) report normalized measures of science test score, share of rote learning score, share of application-based learning score, and creativity score, respectively (mean zero and sd one in the control group). All specifications include the baseline value of the corresponding dependent variable, block fixed effects corresponding to the stratified randomization design, and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.8: Impact of the Lab on Bike on Non-cognitive Measures

	Friends	Extraversion	Agreeableness	Conscientiousness	Emotional stability	Openness
	(1)	(2)	(3)	(4)	(5)	(6)
LoB	-0.007	-0.127	-0.193**	-0.130	-0.135	-0.167
	(0.156)	(0.125)	(0.087)	(0.122)	(0.126)	(0.104)
Control mean	4.07	4.68	4.70	4.78	4.81	4.58
Observations	7,521	7,561	7,560	7,557	7,558	7,559
R-squared	0.037	0.025	0.033	0.039	0.047	0.039
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76	76

Notes: This table reports the impact of the Lab on Bike (LoB) treatment on non-cognitive outcomes using endline survey data. All dependent variables are measured on their original scales. Column (1) reports the number of friends. Columns (2)–(6) report measures of the Big Five personality traits: extraversion, agreeableness, conscientiousness, emotional stability, and openness. The explanatory variable is a binary indicator for schools assigned to receive the LoB treatment. All specifications include block fixed effects corresponding to the stratified randomization design and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels based on unadjusted t-tests.

Table A.9: Did the Mobile Science Lab Change Teachers' Instructional Practices?

	Instill belief	Instill motivation	Instill critical thinking	Use science experiments	Use other examples	Use digital
	(1)	(2)	(3)	(4)	(5)	(6)
MSL	-0.0795	0.0541	0.266	-0.157	-0.246	0.103
	(0.232)	(0.243)	(0.208)	(0.257)	(0.236)	(0.188)
Control mean	0.61	0.50	0.22	0.39	0.33	0.17
Observations	45	45	45	45	45	45
R^2	0.395	0.268	0.610	0.455	0.527	0.562
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
School-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	43	43	43	43	43	43

Notes: This table reports the impact of the Mobile Science Lab (MSL) intervention on science teachers' pedagogical practices. The unit of observation is the science teacher. The key independent variable is an indicator for schools assigned to receive the MSL treatment, estimated using specification 1. All regressions include block fixed effects corresponding to the randomization strata and baseline school-level controls. We further control for teachers characteristics: dummy for a post graduation, dummy for whether science was major in their highest education, caste, gender, age and experience. Robust standard errors clustered at the school level are reported in parentheses. Dependent variables measure teacher-reported instructional practices: instilling belief in students' abilities (column 1), instilling motivation (column 2), promoting critical thinking (column 3), use of science experiments (column 4), use of alternative examples to clarify concepts (column 5), and use of digital equipment in instruction (column 6). Control means are reported for reference. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.10: Did the Lab on Bike Change Teachers' Instructional Practices?

	Instill belief	Instill motivation	Instill critical thinking	Use science experiments	Use other examples	Use digital
	(1)	(2)	(3)	(4)	(5)	(6)
LoB	-0.304	0.122	0.0453	-0.0681	-0.00239	0.175
	(0.200)	(0.231)	(0.213)	(0.167)	(0.237)	(0.146)
Control mean	0.39	0.22	0.22	0.26	0.22	0.09
Observations	50	50	50	50	50	50
R^2	0.430	0.420	0.540	0.579	0.266	0.673
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
School-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	48	48	48	48	48	48

Notes: This table reports the impact of the Lab on Bike (LoB) intervention on science teachers' pedagogical practices. The unit of observation is the science teacher. The key independent variable is an indicator for schools assigned to receive the LoB treatment, estimated using specification 1. All regressions include block fixed effects corresponding to the randomization strata and baseline school-level controls. We further control for teachers characteristics: dummy for a post graduation, dummy for whether science was major in their highest education, caste, gender, age and experience. Robust standard errors clustered at the school level are reported in parentheses. Dependent variables measure teacher-reported instructional practices: instilling belief in students' abilities (column 1), instilling motivation (column 2), promoting critical thinking (column 3), use of science experiments (column 4), use of alternative examples to clarify concepts (column 5), and use of digital equipment in instruction (column 6). Control means are reported for reference. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.11: Dose-Response Effects of Lab on Bike on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Education Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
LoB × Sessions = 0	-0.026	0.071	0.145**	-0.136*	-0.129*
	(0.056)	(0.058)	(0.058)	(0.072)	(0.074)
LoB × Sessions = 1	-0.071	-0.018	0.076	-0.007	-0.127**
	(0.077)	(0.075)	(0.049)	(0.063)	(0.054)
LoB × Sessions = 2	-0.097	-0.033	0.100*	-0.077	-0.145**
	(0.079)	(0.082)	(0.055)	(0.079)	(0.062)
LoB × Sessions = 3	-0.173**	-0.152*	0.154***	-0.030	-0.095
	(0.076)	(0.076)	(0.053)	(0.056)	(0.059)
LoB × Sessions = 4	-0.079	-0.022	0.020	0.034	-0.016
	(0.071)	(0.085)	(0.068)	(0.069)	(0.074)
LoB × Sessions = 5	0.018	0.043	-0.083	0.144**	0.083
	(0.071)	(0.082)	(0.054)	(0.058)	(0.061)
LoB × Sessions = 6	0.183***	0.210***	-0.089	0.121***	0.055
	(0.063)	(0.065)	(0.057)	(0.045)	(0.056)
Control mean	0.00	0.00	0.00	0.00	0.00
Observations	8,453	8,453	8,453	7,708	7,069
R^2	0.071	0.066	0.052	0.046	0.028
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76

Notes: This table reports heterogeneous treatment effects of the Learning on Bus (LoB) intervention by student participation intensity. The explanatory variables are indicators for the number of sessions attended interacted with the LoB treatment assignment, following specification 2. The omitted category is students in control schools. Columns (1)–(5) report effects on the curiosity index, self-confidence index, self-perceived barriers to education index, educational aspiration to study in the future, and self-efficacy, respectively. All specifications include block fixed effects corresponding to the randomization strata, school-level baseline controls, and student-level controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.12: Dose–Response Effects of the Lab on Bike on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
LoB × Sessions = 0	-0.056	-0.035	-0.102*	-0.143**	-0.031	-0.120
	(0.069)	(0.056)	(0.058)	(0.067)	(0.053)	(0.161)
LoB × Sessions = 1	0.035	-0.149*	-0.061	0.030	-0.068	-0.101
	(0.075)	(0.080)	(0.072)	(0.066)	(0.069)	(0.151)
LoB × Sessions = 2	-0.088	-0.158**	-0.138**	-0.111	-0.131	0.158
	(0.064)	(0.073)	(0.069)	(0.074)	(0.082)	(0.141)
LoB × Sessions = 3	-0.115	-0.173**	-0.215***	-0.106	-0.187**	0.005
	(0.070)	(0.071)	(0.074)	(0.068)	(0.073)	(0.151)
LoB × Sessions = 4	-0.007	-0.056	-0.112	-0.099	-0.145*	-0.233
	(0.070)	(0.069)	(0.073)	(0.085)	(0.073)	(0.173)
LoB × Sessions = 5	0.041	-0.071	-0.017	-0.002	-0.064	-0.162
	(0.072)	(0.086)	(0.070)	(0.070)	(0.080)	(0.143)
LoB × Sessions = 6	0.237***	0.100	0.102	0.203***	0.189***	-0.061
	(0.070)	(0.061)	(0.064)	(0.061)	(0.056)	(0.137)
Control mean	-0.00	-0.00	0.00	0.00	-0.00	3.31
Observations	8,453	8,453	8,453	7,433	8,453	7,331
R^2	0.061	0.074	0.069	0.068	0.093	0.032
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76	76

Notes: This table reports heterogeneous treatment effects of the Learning on Bus (LoB) intervention by student participation intensity. The explanatory variables are indicators for the number of sessions attended interacted with the LoB treatment assignment, following specification 2. The omitted category is students in control schools. Columns (1)–(6) report effects on the science interest index, perceived performance in science index, understanding of the science teacher index, affective engagement (enjoy school), consolidated engagement index, and reported school attendance, respectively. All specifications include block fixed effects corresponding to the randomization strata, school-level baseline controls, and student-level controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.13: Dose–Response Effects of the Lab on Bike on Student Performance

	Science Choice	Test Score	Rote Score	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
LoB × Sessions = 0	-0.063**	-0.231**	-0.173**	-0.226**	-0.370***
	(0.030)	(0.098)	(0.083)	(0.089)	(0.101)
LoB × Sessions = 1	0.031	0.066	0.075	0.032	-0.112
	(0.031)	(0.085)	(0.071)	(0.080)	(0.102)
LoB × Sessions = 2	-0.037	-0.088	-0.023	-0.141*	-0.107
	(0.033)	(0.090)	(0.076)	(0.081)	(0.111)
LoB × Sessions = 3	-0.001	0.065	0.044	0.038	-0.135
	(0.032)	(0.106)	(0.087)	(0.100)	(0.111)
LoB × Sessions = 4	-0.027	-0.078	-0.024	-0.122	-0.222**
	(0.037)	(0.106)	(0.092)	(0.095)	(0.109)
LoB × Sessions = 5	0.042	0.028	0.022	-0.009	-0.100
	(0.036)	(0.108)	(0.105)	(0.099)	(0.121)
LoB × Sessions = 6	0.128***	0.229**	0.128	0.216**	-0.017
	(0.036)	(0.107)	(0.085)	(0.102)	(0.111)
Control mean	0.36	0.00	0.00	0.00	0.00
Observations	8,453	8,453	8,453	8,453	7,641
R^2	0.032	0.098	0.077	0.064	0.138
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76

Notes: This table reports heterogeneous treatment effects of the Learning on Bus (LoB) intervention by student participation intensity. The explanatory variables are indicators for the number of sessions attended interacted with the LoB treatment assignment, following specification 2. The omitted category is students in control schools. Column (1) reports effects on an indicator for choosing the science stream in the future. Column (2) reports effects on the survey-based science test score. Columns (3) and (4) report effects on the rote and application-based components of the test, respectively. Column (5) reports effects on the creativity score. All specifications include block fixed effects corresponding to the randomization strata, school-level baseline controls, and student-level controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.14: Heterogeneous Effects by Student-Liked Mobile Science Lab Components on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Education Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
MSL × Science experiment	0.287***	0.194***	-0.246***	0.270***	0.197***
	(0.051)	(0.055)	(0.057)	(0.048)	(0.062)
MSL × Instructional style	0.231***	0.145**	-0.324***	0.076	0.039
	(0.061)	(0.067)	(0.067)	(0.056)	(0.074)
MSL × Group activity	0.060	0.013	-0.110	0.056	0.118
	(0.079)	(0.079)	(0.080)	(0.064)	(0.088)
MSL × Real life examples	-0.121*	-0.025	0.066	-0.058	-0.179*
	(0.064)	(0.067)	(0.078)	(0.077)	(0.100)
MSL × Freedom to express	0.017	0.014	-0.057	0.002	0.060
	(0.077)	(0.067)	(0.076)	(0.061)	(0.056)
MSL × Like nothing	0.135	0.224	-0.154*	-0.030	0.120
	(0.116)	(0.155)	(0.080)	(0.083)	(0.118)
Control mean	-0.00	0.00	-0.00	-0.00	0.00
Observations	5,523	5,523	5,523	5,202	4,832
R^2	0.045	0.034	0.039	0.049	0.021
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects by the components of the Mobile Science Lab (MSL) intervention that students reported liking. The explanatory variables interact the MSL treatment indicator with indicators for the specific components students reported enjoying: science experiments, instructional style, group activity, real-life examples, freedom to express, and like nothing, following specification 2. The omitted category is students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(5) report treatment effects on the curiosity index, self-confidence index, self-perceived education barriers index, educational aspiration to study further, and self-efficacy, respectively. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.15: Heterogeneous Effects by Student-Liked Mobile Science Lab Components on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
MSL × Science experiment	0.297***	0.216***	0.242***	0.300***	0.290***	0.127
	(0.050)	(0.055)	(0.053)	(0.046)	(0.054)	(0.097)
MSL × Instructional style	0.237***	0.165**	0.211***	0.297***	0.236***	-0.119
	(0.050)	(0.080)	(0.071)	(0.043)	(0.066)	(0.155)
MSL × Group activity	0.126	0.097	0.075	0.024	0.051	-0.379***
	(0.083)	(0.076)	(0.070)	(0.074)	(0.092)	(0.122)
MSL × Real life examples	-0.105	-0.055	-0.119	-0.115*	-0.152**	-0.138
	(0.068)	(0.064)	(0.074)	(0.059)	(0.068)	(0.155)
MSL × Freedom to express	0.068	-0.029	0.020	0.026	0.060	-0.079
	(0.066)	(0.086)	(0.087)	(0.065)	(0.066)	(0.126)
MSL × Liked nothing	0.087	0.019	0.229**	0.103	0.139	-0.128
	(0.126)	(0.141)	(0.090)	(0.123)	(0.122)	(0.242)
Control mean	0.00	0.00	-0.00	-0.00	-0.00	3.34
Observations	5,523	5,523	5,523	5,147	5,523	4,855
R^2	0.047	0.039	0.047	0.044	0.047	0.026
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects by the components of the Mobile Science Lab (MSL) intervention that students reported liking on engagement-related outcomes. The explanatory variables interact the MSL treatment indicator with indicators for the specific components students reported enjoying— science experiments, instructional style, group activity, real-life examples, freedom to express, and like nothing, following specification 2. The omitted category is students in control schools. All dependent variables (except the reported attendance) are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(6) report treatment effects on the science interest index, self-perceived performance in science index, understanding of the science teacher index, affective engagement (enjoy school), composite engagement index, and self-reported attendance, respectively. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.16: Heterogeneous Effects by Student-Liked Mobile Science Lab Components on Student Performance

	Science Choice	Test score	Rote	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
MSL × Science Experiment	0.091***	0.184*	0.238**	0.083	0.066
	(0.027)	(0.096)	(0.098)	(0.075)	(0.067)
MSL × Instructional style	0.018	0.185**	0.147**	0.164*	0.028
	(0.037)	(0.084)	(0.070)	(0.091)	(0.062)
MSL × Group activity	0.051	0.024	0.055	-0.020	0.038
	(0.037)	(0.089)	(0.080)	(0.100)	(0.060)
MSL × Real life examples	-0.168***	-0.147*	-0.136*	-0.113	-0.271***
	(0.032)	(0.084)	(0.073)	(0.088)	(0.064)
MSL × Freedom to express	-0.009	-0.027	-0.105	0.063	-0.089
	(0.027)	(0.076)	(0.079)	(0.070)	(0.071)
MSL × Like nothing	-0.110**	-0.170	-0.137	-0.109	-0.216**
	(0.052)	(0.132)	(0.110)	(0.129)	(0.105)
Control mean	0.00	0.00	0.00	-0.00	-0.00
Observations	5,523	5,523	5,523	5,523	5,154
R^2	0.026	0.063	0.080	0.037	0.282
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects by the components of the Mobile Science Lab (MSL) intervention that students reported liking on test-based outcomes. The explanatory variables interact the MSL treatment indicator with indicators for the specific components students reported enjoying—science experiments, instructional style, group activity, real-life examples, freedom to express, and like nothing, following specification 2. The omitted category is students in control schools. Columns (1)–(5) report treatment effects on an indicator for choosing the science stream in the future, the survey-based science test score, the rote component of the test, the application-based component of the test, and the creativity score, respectively. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.17: Heterogeneous Effects of the Mobile Science Lab by Household Wealth on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Education Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
MSL	0.186**	0.171*	-0.261***	0.158***	0.112*
	(0.078)	(0.086)	(0.068)	(0.054)	(0.056)
Low asset dummy	-0.026	-0.001	0.061	-0.084*	-0.071
	(0.035)	(0.040)	(0.047)	(0.045)	(0.044)
MSL × Low asset dummy	-0.021	-0.040	0.085	-0.039	-0.089
	(0.049)	(0.058)	(0.063)	(0.056)	(0.063)
Control mean	0.00	0.00	0.00	0.01	0.00
Observations	5,750	5,750	5,750	5,386	4,994
R^2	0.029	0.027	0.030	0.039	0.014
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention by household wealth status. The main explanatory variable interacts the MSL treatment indicator with a low-asset dummy (equal to one for households below the median wealth level), following specification 3. The omitted category is students in control schools from above-median wealth households. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(5) report treatment effects on the curiosity index, self-confidence index, self-perceived education barriers index, goals and aspirations to study further, and self-efficacy, respectively. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.18: Heterogeneous Effects of the Mobile Science Lab by Household Wealth on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
MSL	0.191***	0.167*	0.176**	0.190***	0.192**	-0.074
	(0.065)	(0.093)	(0.082)	(0.065)	(0.082)	(0.124)
Low asset dummy	-0.086**	-0.029	-0.051	-0.071*	-0.005	0.012
	(0.039)	(0.036)	(0.045)	(0.042)	(0.037)	(0.092)
MSL × Low asset dummy	-0.029	-0.060	-0.060	-0.022	-0.061	-0.082
	(0.060)	(0.051)	(0.062)	(0.056)	(0.049)	(0.124)
Control mean	0.00	0.00	-0.00	-0.00	-0.00	3.34
Observations	5,750	5,750	5,750	5,296	5,750	5,019
R^2	0.033	0.030	0.035	0.024	0.029	0.021
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention on engagement-related outcomes by household wealth status. The explanatory variables include the MSL treatment indicator, a low-asset dummy (below median household wealth), and their interaction, following specification 3. The omitted category is students in control schools from above-median wealth households. All dependent variables (except reported attendance) are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(6) report treatment effects on the science interest index, self-perceived performance in science index, understanding of the science teacher index, affective engagement (enjoy school), composite engagement index, and self-reported attendance, respectively. All specifications include block fixed effects and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.19: Heterogeneous Effects of the Mobile Science Lab by Household Wealth on Student Performance

	Science Choice	Test score	Rote	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
MSL	0.013	-0.018	0.051	-0.057	-0.083
	(0.038)	(0.123)	(0.115)	(0.102)	(0.096)
Low Asset	-0.027	-0.199***	-0.168***	-0.147***	-0.127***
	(0.021)	(0.045)	(0.042)	(0.048)	(0.042)
MSL × Low Asset	-0.006	0.061	0.015	0.062	0.052
	(0.033)	(0.070)	(0.059)	(0.076)	(0.057)
Control mean	0.37	0.00	-0.00	-0.00	-0.00
Observations	5,750	5,750	5,750	5,750	5,326
R^2	0.010	0.059	0.073	0.034	0.283
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention on student performance outcomes by household wealth status. The explanatory variables include the MSL treatment indicator, a low-asset dummy (equal to one for households below the median wealth level), and their interaction, following specification 3. The omitted category is students in control schools from above-median wealth households. Column (1) reports effects on an indicator for choosing the science stream in the future. Column (2) reports effects on the survey-based science test score. Columns (3) and (4) report effects on the rote and application-based components of the test, respectively. Column (5) reports effects on the creativity score. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.20: Heterogeneous Effects of the Mobile Science Lab by Gender on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Education Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
MSL	0.166**	0.160*	-0.245***	0.214***	0.149**
	(0.080)	(0.082)	(0.074)	(0.072)	(0.065)
Female	0.058	0.066	-0.024	0.137***	0.053
	(0.057)	(0.062)	(0.053)	(0.049)	(0.045)
MSL × Female	0.021	-0.009	0.025	-0.131*	-0.125*
	(0.080)	(0.078)	(0.072)	(0.072)	(0.072)
Control mean	0.00	0.00	0.00	0.01	0.00
Observations	5,750	5,750	5,750	5,386	4,994
R^2	0.030	0.027	0.027	0.039	0.012
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention by gender. The explanatory variables include the MSL treatment indicator, a female dummy, and their interaction, following specification 3. The omitted category is male students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(5) report treatment effects on the curiosity index, self-confidence index, self-perceived education barriers index, educational aspirations to study further, and self-efficacy, respectively. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.21: Heterogeneous Effects of the Mobile Science Lab by Gender on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
MSL	0.197**	0.146	0.138*	0.179**	0.171**	-0.048
	(0.075)	(0.087)	(0.081)	(0.072)	(0.078)	(0.139)
Female	0.062	0.101	0.067	0.104*	0.070	0.071
	(0.053)	(0.064)	(0.052)	(0.056)	(0.060)	(0.064)
MSL × Female	-0.030	-0.002	0.028	0.004	-0.002	-0.102
	(0.068)	(0.077)	(0.078)	(0.071)	(0.076)	(0.115)
Control mean	0.00	0.00	-0.00	-0.00	-0.00	3.34
Observations	5,750	5,750	5,750	5,296	5,750	5,019
R^2	0.031	0.032	0.035	0.025	0.029	0.021
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention on engagement-related outcomes by gender. The explanatory variables include the MSL treatment indicator, a female dummy, and their interaction, following specification 3. The omitted category is male students in control schools. All dependent variables (except attendance) are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(6) report treatment effects on the science interest index, self-perceived performance in science index, understanding of the science teacher index, affective engagement (enjoy school), composite engagement index, and self-reported attendance. All specifications include block fixed effects and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.22: Heterogeneous Effects of the Mobile Science Lab by Gender on Student Performance

	Science Choice	Test score	Rote	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
MSL	0.001	0.088	0.133	0.035	-0.005
	(0.036)	(0.119)	(0.116)	(0.093)	(0.101)
Female	-0.004	0.139***	0.131***	0.107**	0.127**
	(0.017)	(0.051)	(0.045)	(0.048)	(0.053)
MSL × Female	0.019	-0.158**	-0.143*	-0.132*	-0.111
	(0.028)	(0.074)	(0.072)	(0.076)	(0.075)
Control mean	0.37	0.00	-0.00	-0.00	-0.00
Observations	5,750	5,750	5,750	5,750	5,326
R^2	0.009	0.054	0.069	0.032	0.282
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention on student performance outcomes by gender. The explanatory variables include the MSL treatment indicator, a female dummy, and their interaction, following specification 3. The omitted category is male students in control schools. Column (1) reports effects on an indicator for choosing the science stream in the future. Column (2) reports effects on the survey-based science test score. Columns (3) and (4) report effects on the rote and application-based components of the test, respectively. Column (5) reports effects on the creativity score. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.23: Heterogeneous Effects of the Mobile Science Lab by Student Ability on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Education Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
MSL	0.249***	0.226***	-0.228***	0.168***	0.070
	(0.076)	(0.083)	(0.069)	(0.058)	(0.065)
Low ability dummy	-0.070	-0.029	0.081	-0.090*	-0.074
	(0.065)	(0.072)	(0.051)	(0.045)	(0.047)
MSL × Low ability dummy	-0.099	-0.130	0.013	-0.054	0.064
	(0.089)	(0.087)	(0.081)	(0.068)	(0.090)
Control mean	0.00	0.00	0.00	0.01	0.00
Observations	4,518	4,518	4,518	4,269	3,981
R^2	0.041	0.032	0.029	0.042	0.015
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention by student ability. The explanatory variables include the MSL treatment indicator, a low-ability dummy (equal to one for students scoring below the median on the pre-intervention survey-based test), and their interaction, following specification 3. The omitted category is higher-ability students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(5) report treatment effects on the curiosity index, self-confidence index, self-perceived education barriers index, educational aspirations to study further, and self-efficacy, respectively. All specifications include block fixed effects and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.24: Heterogeneous Effects of the Mobile Science Lab by Student Ability on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
MSL	0.215***	0.223**	0.191**	0.274***	0.242***	0.023
	(0.076)	(0.089)	(0.083)	(0.073)	(0.086)	(0.131)
Low ability dummy	-0.108	-0.122*	-0.149**	-0.048	-0.050	-0.026
	(0.073)	(0.068)	(0.068)	(0.059)	(0.066)	(0.114)
MSL × Low ability dummy	-0.011	-0.101	-0.046	-0.121	-0.107	-0.239
	(0.111)	(0.093)	(0.096)	(0.082)	(0.092)	(0.152)
Control mean	-0.00	-0.00	0.01	0.00	0.01	3.37
Observations	4,518	4,518	4,518	4,249	4,518	4,005
R^2	0.039	0.046	0.047	0.033	0.036	0.023
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention on engagement-related outcomes by student ability. The explanatory variables include the MSL treatment indicator, a low-ability dummy (equal to one for students scoring below the median on the pre-intervention survey-based test), and their interaction, following specification 3. The omitted category is higher-ability students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(6) report treatment effects on the science interest index, self-perceived performance in science index, understanding of the science teacher index, affective engagement (enjoy school), composite engagement index, and self-reported attendance. All specifications include block fixed effects and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.25: Heterogeneous Effects of the Mobile Science Lab by Student Ability on Student Performance

	Science Choice	Test score	Rote	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
MSL	0.047	0.037	0.122	-0.050	-0.131
	(0.039)	(0.152)	(0.140)	(0.123)	(0.127)
Low Ability	-0.035	-0.143	-0.097	-0.148*	-0.133
	(0.032)	(0.088)	(0.073)	(0.082)	(0.085)
MSL × Low Ability	-0.011	-0.090	-0.149	0.014	0.054
	(0.044)	(0.125)	(0.117)	(0.114)	(0.111)
Control mean	0.36	0.04	0.04	0.02	0.03
Observations	4,518	4,518	4,518	4,518	4,518
R^2	0.011	0.065	0.082	0.035	0.319
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention on student performance outcomes by baseline ability. The explanatory variables include the MSL treatment indicator, a low-ability dummy (equal to one for students scoring below the median on the pre-intervention survey-based test), and their interaction, following specification 3. The omitted category is higher-ability students in control schools. Column (1) reports effects on an indicator for choosing the science stream in the future. Column (2) reports effects on the survey-based science test score. Columns (3) and (4) report effects on the rote and application-based components of the test, respectively. Column (5) reports effects on the creativity score. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.26: Heterogeneous Effects by Student-Liked Lab on Bike Components on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
LoB × Science experiment	0.131***	0.046	-0.115***	0.166***	0.060
	(0.046)	(0.050)	(0.036)	(0.036)	(0.041)
LoB × Instructional style	0.285***	0.362***	-0.105**	0.171***	0.015
	(0.045)	(0.052)	(0.051)	(0.041)	(0.059)
LoB × Group activity	0.109	0.131*	-0.087	0.079	0.094
	(0.068)	(0.074)	(0.059)	(0.059)	(0.063)
LoB × Real-life examples	-0.194***	-0.103*	0.164***	-0.216***	-0.199***
	(0.055)	(0.054)	(0.059)	(0.060)	(0.060)
LoB × Freedom to express	0.032	0.023	-0.092**	-0.004	0.029
	(0.037)	(0.043)	(0.044)	(0.046)	(0.039)
LoB × Like nothing	0.092	0.270**	0.059	-0.057	-0.105
	(0.122)	(0.114)	(0.082)	(0.074)	(0.079)
Control mean	0.00	0.00	0.00	0.00	0.00
Observations	8,015	8,015	8,015	7,352	6,755
R^2	0.027	0.035	0.026	0.023	0.009
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76

Notes: This table reports heterogeneous treatment effects by the components of the Lab on Bike (LoB) intervention that students reported liking. The explanatory variables interact the LoB treatment indicator with indicators for the specific components students reported enjoying: science experiments, instructional style, group activity, real-life examples, freedom to express, and like nothing, following specification 2. The omitted category is students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(5) report treatment effects on curiosity, self-confidence, perceived education barriers, goals and aspirations to study further, and self-efficacy, respectively. All specifications include block fixed effects and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.27: Heterogeneous Effects by Student-Liked Lab on Bike on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
LoB × Science experiment	0.215***	0.045	0.090*	0.127***	0.106**	0.084
	(0.048)	(0.050)	(0.048)	(0.045)	(0.045)	(0.095)
LoB × Instructional style	0.210***	0.249***	0.231***	0.206***	0.290***	0.087
	(0.044)	(0.043)	(0.044)	(0.050)	(0.047)	(0.083)
LoB × Group activity	0.088	0.139**	0.078	0.084	0.058	0.252**
	(0.060)	(0.067)	(0.059)	(0.053)	(0.057)	(0.118)
LoB × Real-life examples	-0.218***	-0.167***	-0.277***	-0.254***	-0.278***	-0.170
	(0.059)	(0.054)	(0.060)	(0.048)	(0.054)	(0.125)
LoB × Freedom to express	-0.011	-0.016	-0.014	0.039	-0.005	-0.109
	(0.056)	(0.045)	(0.045)	(0.043)	(0.044)	(0.112)
LoB × Like nothing	-0.047	0.064	-0.039	-0.031	0.020	-0.095
	(0.093)	(0.110)	(0.098)	(0.107)	(0.090)	(0.147)
Control mean	-0.00	-0.00	0.00	0.00	-0.00	3.31
Observations	8,015	8,015	8,015	7,138	8,015	6,986
R^2	0.027	0.030	0.026	0.026	0.033	0.031
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76 76	76	

Notes: This table reports heterogeneous treatment effects by the components of the Lab on Bike (LoB) intervention that students reported liking. The explanatory variables interact the LoB treatment indicator with indicators for the specific components students reported enjoying: science experiments, instructional style, group activity, real-life examples, freedom to express, and like nothing, following specification 2. The omitted category is students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(6) report treatment effects on the science interest index, self-perceived performance in science index, understanding of the science teacher index, affective engagement (enjoy school), composite engagement index, and self-reported attendance. All specifications include block fixed effects and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.28: Heterogeneous Effects by Student-Liked Lab on Bike on Student Performance

	Science choice	Test score	Rote	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
LoB × Science experiment	0.127***	0.131*	0.119**	0.068	0.047
	(0.024)	(0.071)	(0.058)	(0.069)	(0.072)
LoB × Instructional style	-0.019	0.097	0.032	0.119*	-0.112*
	(0.025)	(0.064)	(0.055)	(0.068)	(0.060)
LoB × Group activity	0.057**	0.242***	0.232***	0.169**	0.050
	(0.026)	(0.068)	(0.069)	(0.067)	(0.078)
LoB × Real-life examples	-0.076***	-0.065	-0.031	-0.070	-0.203***
	(0.028)	(0.069)	(0.064)	(0.069)	(0.068)
LoB × Freedom to express	-0.011	0.098	0.075	0.074	-0.038
	(0.024)	(0.067)	(0.060)	(0.068)	(0.072)
LoB × Like nothing	-0.114***	-0.094	-0.107	-0.031	-0.309***
	(0.037)	(0.108)	(0.089)	(0.111)	(0.092)
Control mean	0.36	0.00	-0.00	0.00	0.00
Observations	8,015	8,015	8,015	8,015	7,297
R^2	0.025	0.079	0.060	0.048	0.112
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76

Notes: This table reports heterogeneous treatment effects by the components of the Lab on Bike (LoB) intervention that students reported liking on test-based outcomes. The explanatory variables interact the LoB treatment indicator with indicators for the specific components students reported enjoying—science experiments, instructional style, group activity, real-life examples, freedom to express, and like nothing, following specification 2. The omitted category is students in control schools. Columns (1)–(5) report treatment effects on an indicator for choosing the science stream in the future, the survey-based science test score, the rote component of the test, the application-based component of the test, and the creativity score, respectively. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.29: Heterogeneous Effects of the Lab-on-Bike by Household Wealth on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Education Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
LoB	-0.044	0.023	0.063	-0.002	-0.033
	(0.055)	(0.062)	(0.039)	(0.047)	(0.053)
Low asset dummy	-0.148***	-0.053	0.105**	-0.140***	-0.049
	(0.037)	(0.034)	(0.042)	(0.035)	(0.042)
LoB × Low asset dummy	0.059	-0.004	-0.041	0.040	-0.050
	(0.054)	(0.046)	(0.055)	(0.055)	(0.055)
Control mean	0.00	0.00	0.00	0.00	0.00
Observations	8,453	8,453	8,453	7,708	7,069
R^2	0.017	0.019	0.021	0.014	0.006
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table reports heterogeneous treatment effects of the Lab-on-Bike (LoB) intervention by household wealth status. The main explanatory variable interacts the LoB treatment indicator with a low-asset dummy (equal to one for households below the median wealth level), following specification 3. The omitted category is students in control schools from above-median wealth households. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(5) report treatment effects on the curiosity index, self-confidence index, self-perceived education barriers index, goals and aspirations to study further, and self-efficacy, respectively. All specifications include block fixed effects and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.30: Heterogeneous Effects of the Lab-on-Bike by Household Wealth on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
LoB	0.000	-0.086	-0.090	-0.024	-0.068	-0.060
	(0.060)	(0.054)	(0.056)	(0.054)	(0.057)	(0.129)
Low asset dummy	-0.119***	-0.104**	-0.137***	-0.110**	-0.097***	-0.025
	(0.036)	(0.040)	(0.041)	(0.043)	(0.035)	(0.065)
LoB × Low asset dummy	0.051	0.040	0.052	0.035	0.056	-0.027
	(0.053)	(0.052)	(0.055)	(0.061)	(0.052)	(0.095)
Control mean	0.00	0.00	0.00	0.00	0.00	3.31
Observations	8,453	8,453	8,453	7,433	8,453	7,331
R^2	0.012	0.022	0.017	0.015	0.019	0.027
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76	76

Notes: This table reports heterogeneous treatment effects of the Lab-on-Bike (LoB) intervention by household wealth status. The main explanatory variable interacts the LoB treatment indicator with a low-asset dummy (equal to one for households below the median wealth level), following specification 3. The omitted category is students in control schools from above-median wealth households. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(6) report treatment effects on the science interest index, self-perceived performance in science index, understanding of the science teacher index, affective engagement (enjoy school), composite engagement index, and self-reported attendance, respectively. All specifications include block fixed effects and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.31: Heterogeneous Effects of the Lab-on-Bike by Household Wealth on Student Performance

	Science Choice	Test Score	Rote Score	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
LoB	0.021	-0.037	-0.028	-0.060	-0.166
	(0.028)	(0.091)	(0.070)	(0.083)	(0.104)
Low asset dummy	-0.068***	-0.224***	-0.170***	-0.196***	-0.182***
	(0.018)	(0.043)	(0.031)	(0.046)	(0.046)
LoB × Low asset dummy	0.007	0.186***	0.161***	0.165**	0.084
	(0.028)	(0.060)	(0.050)	(0.065)	(0.058)
Control mean	0.36	0.00	-0.00	0.00	0.00
Observations	8,453	8,453	8,453	8,453	7,641
R^2	0.012	0.075	0.058	0.047	0.113
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: Robust standard errors clustered at the school level are reported in parentheses. The main explanatory variable interacts the Lab-on-Bike (LoB) treatment indicator with a low-asset dummy (equal to one for households below the median wealth level), following specification 3. The omitted category is students in control schools from above-median wealth households. Test outcomes are standardized to have mean zero and standard deviation one in the control group. Science choice is a binary indicator equal to one if the student reports intending to pursue the science stream in the future. All specifications include block fixed effects and school-level baseline controls. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.32: Heterogeneous Effects of the Lab-on-Bike by Gender on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Education Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
LoB	-0.040	-0.010	0.106**	-0.006	-0.067
	(0.074)	(0.081)	(0.049)	(0.053)	(0.063)
Female	0.053	0.048	-0.009	0.037	0.010
	(0.054)	(0.055)	(0.043)	(0.051)	(0.048)
LoB × Female	0.039	0.066	-0.114**	0.041	0.036
	(0.073)	(0.080)	(0.056)	(0.068)	(0.069)
Control mean	0.00	0.00	0.00	-0.00	0.00
Observations	8,453	8,453	8,453	7,708	7,069
R^2	0.015	0.020	0.022	0.012	0.005
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76

Notes: Robust standard errors clustered at the school level are reported in parentheses. The main explanatory variable interacts the Lab-on-Bike (LoB) treatment indicator with a female dummy, following specification 3. The omitted category is male students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(5) report treatment effects on the curiosity index, self-confidence index, self-perceived education barriers index, goals and aspirations to study further, and self-efficacy. All specifications include block fixed effects and school-level baseline controls. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.33: Heterogeneous Effects of the Lab-on-Bike by Gender on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
LoB	-0.013	-0.092	-0.089	-0.069	-0.134*	-0.100
	(0.074)	(0.069)	(0.074)	(0.065)	(0.071)	(0.149)
Female	0.015	0.037	0.085	0.052	0.034	0.042
	(0.056)	(0.052)	(0.054)	(0.057)	(0.064)	(0.116)
LoB × Female	0.063	0.042	0.040	0.113*	0.168**	0.062
	(0.074)	(0.070)	(0.074)	(0.067)	(0.079)	(0.144)
Control mean	0.00	0.00	0.00	0.00	0.00	3.31
Observations	8,453	8,453	8,453	7,433	8,453	7,331
R^2	0.011	0.021	0.017	0.018	0.024	0.027
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76	76

Notes: Robust standard errors clustered at the school level are reported in parentheses. The main explanatory variable interacts the Lab-on-Bike (LoB) treatment indicator with a female dummy, following specification 3. The omitted category is male students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(6) report treatment effects on the science interest index, self-perceived performance in science index, understanding of the science teacher index, affective engagement (enjoy school), composite engagement index, and self-reported attendance, respectively. All specifications include block fixed effects and school-level baseline controls. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.34: Heterogeneous Effects of the Lab-on-Bike by Gender on Student Performance

	Science Choice	Test Score	Rote	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
LoB	0.011	0.013	0.007	-0.008	-0.071
	(0.031)	(0.090)	(0.070)	(0.087)	(0.099)
Female	-0.036	0.068	0.030	0.057	0.172***
	(0.022)	(0.056)	(0.046)	(0.061)	(0.055)
LoB × Female	0.029	0.034	0.044	0.015	-0.120
	(0.030)	(0.069)	(0.058)	(0.076)	(0.075)
Control mean	0.36	0.00	-0.00	0.00	0.00
Observations	8,453	8,453	8,453	8,453	7,641
R^2	0.009	0.071	0.055	0.044	0.113
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76

Notes: Robust standard errors clustered at the school level are reported in parentheses. The main explanatory variable interacts the Lab-on-Bike (LoB) treatment indicator with a female dummy, following specification 3. The omitted category is male students in control schools. Survey-based test outcomes are standardized to have mean zero and standard deviation one in the control group. Science choice is a binary indicator equal to one if the student reports intending to pursue the science stream in the future. All specifications include block fixed effects and school-level baseline controls. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.35: Heterogeneous Effects of the Lab-on-Bike by Student Ability on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Education Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
LoB	-0.074	0.013	0.041	0.014	-0.090
	(0.066)	(0.075)	(0.052)	(0.056)	(0.059)
Low ability	-0.247***	-0.213***	0.016	-0.096**	-0.128***
	(0.064)	(0.063)	(0.049)	(0.037)	(0.047)
LoB × Low ability	0.069	0.060	0.106*	-0.086	0.045
	(0.078)	(0.080)	(0.058)	(0.064)	(0.070)
Control mean	0.06	0.02	-0.06	0.07	0.03
Observations	6,147	6,147	6,147	5,667	5,211
R^2	0.025	0.025	0.023	0.018	0.011
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76

Notes: Robust standard errors clustered at the school level are reported in parentheses. The main explanatory variable interacts the Lab-on-Bike (LoB) treatment indicator with a low-ability dummy (equal to one for students scoring below the median on the baseline survey-based test), following specification 3. The omitted category is high-ability students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(5) report treatment effects on the curiosity index, self-confidence, self-perceived education barriers index, goals and aspirations to study further, and self-efficacy index. All specifications include block fixed effects and school-level baseline controls. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.36: Heterogeneous Effects of the Lab-on-Bike by Student Ability on Engagement Factors

	Science interest	Performance in science	Understanding science teacher	Engagement index
	(1)	(2)	(3)	(4)
LoB	-0.012	-0.101	-0.123*	-0.092
	(0.070)	(0.061)	(0.064)	(0.068)
Low ability	-0.253***	-0.253***	-0.248***	-0.228***
	(0.053)	(0.055)	(0.054)	(0.056)
LoB × Low ability	0.024	0.083	0.080	0.083
	(0.070)	(0.068)	(0.072)	(0.072)
Control mean	0.04	0.03	0.04	0.04
Observations	6,147	6,147	6,147	6,147
R^2	0.030	0.033	0.032	0.027
Block fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76

Notes: Robust standard errors clustered at the school level are reported in parentheses. The main explanatory variable interacts the Lab-on-Bike (LoB) treatment indicator with a low-ability dummy (equal to one for students scoring below the median on the baseline survey-based test), following specification 3. The omitted category is high-ability students in control schools. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(6) report treatment effects on the science interest index, self-perceived performance in science index, understanding of the science teacher index, affective engagement (enjoy school), composite engagement index, and self-reported attendance, respectively. All specifications include block fixed effects and school-level baseline controls. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.37: Heterogeneous Effects of the Lab-on-Bike by Student Ability on Student Performance

	Science Choice	Test Score	Rote	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
LoB	0.030	-0.010	0.023	-0.062	-0.080
	(0.030)	(0.099)	(0.074)	(0.099)	(0.111)
Low ability	-0.028	-0.119*	-0.061	-0.125*	-0.096
	(0.021)	(0.063)	(0.060)	(0.070)	(0.075)
LoB × Low ability	-0.037	-0.080	-0.131*	0.000	-0.178*
	(0.031)	(0.087)	(0.073)	(0.098)	(0.103)
Control mean	0.38	0.10	0.09	0.07	0.02
Observations	6,147	6,147	6,147	6,147	6,147
R^2	0.013	0.077	0.060	0.049	0.130
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76

Notes: Robust standard errors clustered at the school level are reported in parentheses. The main explanatory variable interacts the Lab-on-Bike (LoB) treatment indicator with a low-ability dummy (equal to one for students scoring below the median on the baseline survey-based test), following specification 3. The omitted category is high-ability students in control schools. All survey-based test outcomes are standardized to have mean zero and standard deviation one in the control group. Science choice is a binary indicator equal to one if the student reports intending to pursue the science stream in the future. All specifications include block fixed effects and school-level baseline controls. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.38: Heterogeneous Effects of the Mobile Science Lab by Teachers' Quality on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Education Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
MSL	-0.047	-0.129	-0.026	0.157*	-0.042
	(0.123)	(0.127)	(0.101)	(0.092)	(0.077)
Low Teacher Quality dummy	-0.268	-0.232	0.231	0.029	-0.246**
	(0.199)	(0.217)	(0.149)	(0.116)	(0.111)
MSL × Low Teacher Quality dummy	0.480*	0.599**	-0.446**	-0.045	0.300**
	(0.251)	(0.280)	(0.186)	(0.157)	(0.142)
Control mean	0.06	0.06	0.01	0.01	0.00
Observations	4,256	4,256	4,256	3,997	3,717
R^2	0.042	0.060	0.045	0.043	0.019
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	41	41	41	41	41

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention by teachers' quality. The main explanatory variable interacts the MSL treatment indicator with a teacher quality dummy (equal to one if self-reported teachers' quality is low), following specification 3. The omitted category is students in control schools from high-quality teacher. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(5) report treatment effects on the curiosity index, self-confidence index, self-perceived education barriers index, educational aspirations to study further, and self-efficacy index, respectively. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.39: Heterogeneous Effects of the Mobile Science Lab by Teachers Quality on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
MSL	0.219**	-0.126	-0.052	-0.018	-0.102	-0.252
	(0.098)	(0.140)	(0.128)	(0.108)	(0.138)	(0.192)
Low Teacher Quality dummy	0.180	-0.141	-0.242	-0.224	-0.298	-0.552*
	(0.159)	(0.172)	(0.200)	(0.187)	(0.217)	(0.321)
MSL × Low Teacher Quality dummy	-0.192	0.468*	0.424*	0.415*	0.601**	0.469
	(0.186)	(0.243)	(0.234)	(0.219)	(0.273)	(0.354)
Control mean	0.08	0.08	0.08	0.05	0.05	3.32
Observations	4,256	4,256	4,256	3,942	4,256	3,743
R^2	0.026	0.058	0.047	0.031	0.051	0.035
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	41	41	41	41	41	41

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention by teachers' quality. The main explanatory variable interacts the MSL treatment indicator with a teacher quality dummy (equal to one if self-reported teachers' quality is low), following specification 3. The omitted category is students in control schools from high-quality teacher. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(6) report treatment effects on the science interest index, self-perceived performance in science index, understanding of the science teacher index, affective engagement (enjoy school), composite engagement index, and self-reported attendance, respectively. All specifications include block fixed effects and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.40: Heterogeneous Effects of the Mobile Science Lab by Teachers Quality on Student Performance

	Science Choice	Test score	Rote	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
MSL	0.037	0.012	0.158	-0.110	-0.070
	(0.039)	(0.143)	(0.128)	(0.117)	(0.123)
Low Teacher Quality dummy	0.153***	0.060	0.052	0.026	-0.186
	(0.052)	(0.166)	(0.152)	(0.139)	(0.162)
MSL × Low Teacher Quality dummy	-0.083	0.136	0.168	0.074	0.040
	(0.079)	(0.248)	(0.216)	(0.200)	(0.224)
Control mean	0.40	0.08	-0.01	0.13	0.06
Observations	4,256	4,256	4,256	4,256	4,013
R^2	0.014	0.081	0.082	0.058	0.249
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	41	41	41	41	41

Notes: This table reports heterogeneous treatment effects of the Mobile Science Lab (MSL) intervention by teachers' quality. The main explanatory variable interacts the MSL treatment indicator with a teacher quality dummy (equal to one if self-reported teachers' quality is low), and their interaction, following specification 3. The omitted category is students in control schools from high-quality teacher. Column (1) reports effects on an indicator for choosing the science stream in the future. Column (2) reports effects on the survey-based science test score. Columns (3) and (4) report effects on the rote and application-based components of the test, respectively. Column (5) reports effects on the creativity score. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.41: Heterogeneous Effects of the Lab-on-Bike by Teachers' Quality on Personal Factors

	Curiosity	Self-confidence	Barriers to education	Education Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
LoB	0.057	0.079	0.018	0.021	0.006
	(0.084)	(0.086)	(0.057)	(0.065)	(0.065)
Low Teacher Quality dummy	-0.061	-0.021	0.009	-0.098	0.070
	(0.149)	(0.158)	(0.116)	(0.142)	(0.110)
LoB × Low Teacher Quality dummy	-0.165	-0.115	-0.025	0.101	-0.047
	(0.185)	(0.202)	(0.133)	(0.156)	(0.152)
Control mean	-0.01	0.01	0.02	-0.02	-0.00
Observations	6,537	6,537	6,537	5,902	5,376
R^2	0.023	0.031	0.024	0.016	0.009
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	58	58	58	58	58

Notes: This table reports heterogeneous treatment effects of the Lab-on-Bike (LoB) intervention by teachers' quality. The main explanatory variable interacts the LoB treatment indicator with a teacher quality dummy (equal to one if self-reported teachers' quality is low), following specification 3. The omitted category is students in control schools from high-quality teachers. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(5) report treatment effects on the curiosity index, self-confidence index, self-perceived education barriers index, educational aspirations to study further, and self-efficacy, respectively. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.42: Heterogeneous Effects of the Lab-on-Bike by Teachers Quality on Engagement Factors

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
LoB	0.084	0.027	0.019	0.080	0.054	-0.021
	(0.079)	(0.085)	(0.076)	(0.069)	(0.070)	(0.137)
Low Teacher Quality dummy	-0.310**	-0.057	-0.077	-0.026	0.018	-0.029
	(0.118)	(0.156)	(0.153)	(0.112)	(0.112)	(0.077)
LoB × Low Teacher Quality dummy	0.088	-0.137	-0.105	-0.168	-0.245*	-0.122
	(0.156)	(0.193)	(0.181)	(0.131)	(0.142)	(0.098)
Control mean	0.00	-0.02	-0.00	0.00	0.00	3.29
Observations	6,537	6,537	6,537	5,669	6,537	5,614
R^2	0.025	0.029	0.025	0.025	0.031	0.034
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	58	58	58	58	58	58

Notes: This table reports heterogeneous treatment effects of the Lab-on-Bike (LoB) intervention by teachers' quality. The main explanatory variable interacts the LoB treatment indicator with a teacher quality dummy (equal to one if self-reported teachers' quality is low), following specification 3. The omitted category is students in control schools from high-quality teacher. All dependent variables are standardized indices with mean zero and standard deviation one in the control group. Columns (1)–(6) report treatment effects on the science interest index, self-perceived performance in science index, understanding of the science teacher index, affective engagement (enjoy school), composite engagement index, and self-reported attendance, respectively. All specifications include block fixed effects and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.43: Heterogeneous Effects of the Lab-on-Bike by Teachers Quality on Student Performance

	Science Choice	Test score	Rote	Application	Creativity
	(1)	(2)	(3)	(4)	(5)
LoB	0.027	0.051	0.077	0.002	-0.246**
	(0.035)	(0.139)	(0.111)	(0.120)	(0.111)
Low Teacher Quality dummy	-0.096*	0.022	-0.040	0.116	-0.022
	(0.051)	(0.220)	(0.186)	(0.184)	(0.245)
	(0.052)	(0.166)	(0.152)	(0.139)	(0.162)
LoB × Low Teacher Quality dummy	0.062	-0.158	-0.069	-0.229	0.206
	(0.075)	(0.264)	(0.223)	(0.228)	(0.293)
Control mean	0.36	0.04	0.02	0.04	0.00
Observations	6,537	6,537	6,537	6,537	5,869
R^2	0.013	0.075	0.071	0.043	0.145
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	58	58	58	58	58

Notes: This table reports heterogeneous treatment effects of the Lab-on-Bike (LoB) intervention by teachers' quality. The main explanatory variable interacts the LoB treatment indicator with a teacher quality dummy (equal to one if self-reported teachers' quality is low), and their interaction, following specification 3. The omitted category is students in control schools from high-quality teacher. Column (1) reports effects on an indicator for choosing the science stream in the future. Column (2) reports effects on the survey-based science test score. Columns (3) and (4) report effects on the rote and application-based components of the test, respectively. Column (5) reports effects on the creativity score. All specifications include block fixed effects corresponding to the randomization strata and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table A.44: Impact of Mobile Science Lab on Weekly Study Time

	Science Tuition	Science Home	Mathematics Tuition	Mathematics Home	Other Tuition	Other Home
	(1)	(2)	(3)	(4)	(5)	(6)
MSL	-0.066	0.232*	-0.062	0.095	-0.122	0.180*
	(0.161)	(0.127)	(0.161)	(0.119)	(0.179)	(0.097)
Observations	4,638	4,927	4,484	5,219	4,712	5,082
R^2	0.025	0.013	0.027	0.014	0.028	0.015
Control mean	2.43	2.68	2.43	2.89	2.74	3.42
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55	55

Notes: This table reports the impact of the Mobile Science Lab (MSL) treatment on weekly study time using endline survey data. The dependent variables measure hours devoted per week to studying science, mathematics, and other subjects, separately for tuition and home study. The explanatory variable is a binary indicator for schools assigned to receive the MSL treatment. All specifications include block fixed effects (corresponding to the stratification used in randomization) and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. The sample size varies across columns due to missing responses on specific study-time modules. ***, **, * denote statistical significance at 1, 5, and 10 percent levels based on unadjusted p-values from standard t-tests.

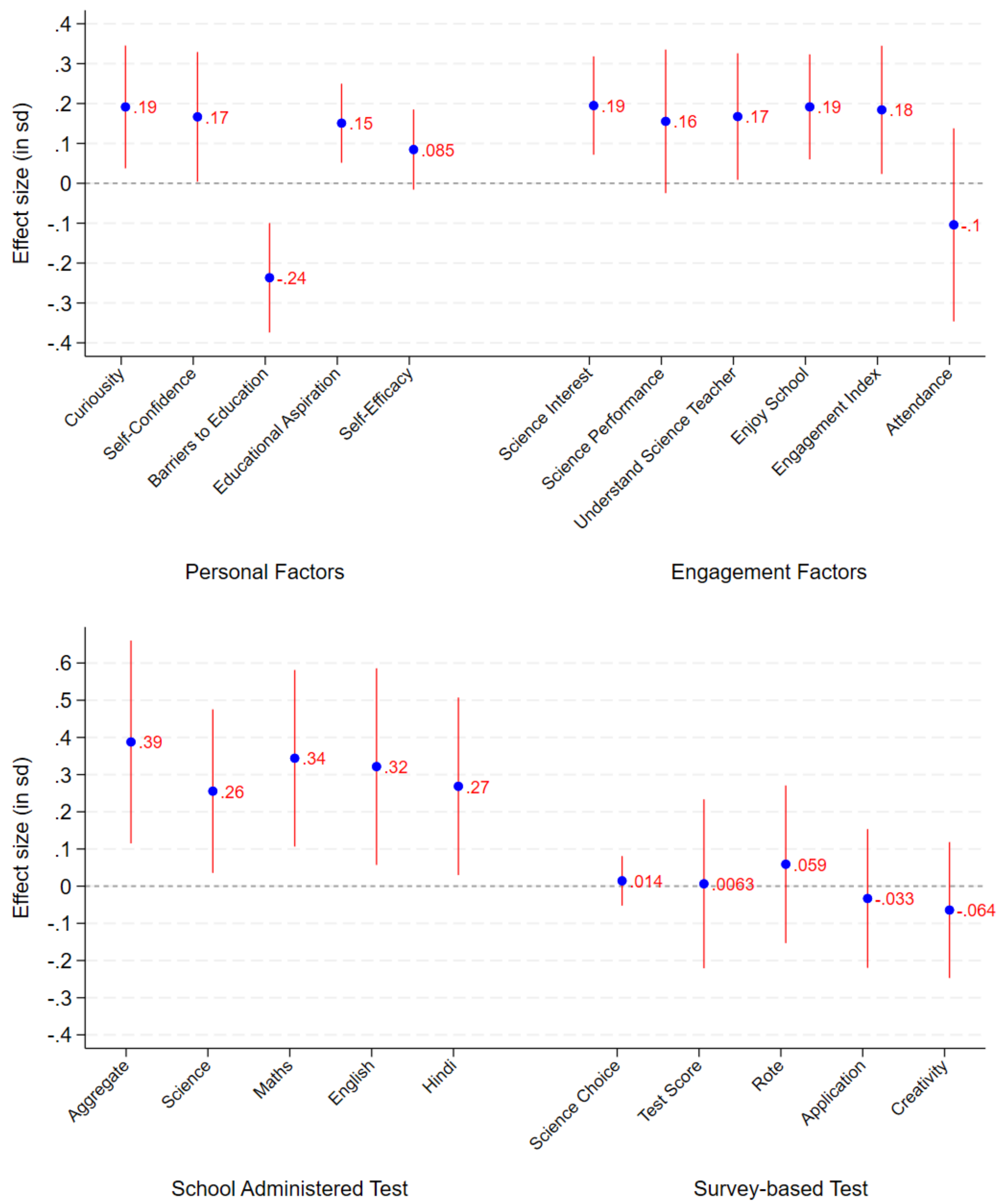
Table A.45: Impact of the Lab on Bike on Weekly Study Time

	Science Tuition	Science Home	Mathematics Tuition	Mathematics Home	Other Tuition	Other Home
	(1)	(2)	(3)	(4)	(5)	(6)
LoB	-0.342**	0.087	-0.231	0.098	-0.352**	0.019
	(0.142)	(0.117)	(0.150)	(0.141)	(0.169)	(0.119)
Observations	6,773	7,053	6,368	7,623	6,828	7,228
R^2	0.091	0.010	0.094	0.019	0.093	0.014
Control mean	2.42	2.82	2.50	3.07	2.76	3.52
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76	76

Notes: This table reports the impact of the Lab on Bike (LoB) treatment on weekly study time using endline survey data. The dependent variables measure hours devoted per week to studying science, mathematics, and other subjects, separately for tuition and home study. The explanatory variable is a binary indicator for schools assigned to receive the LoB treatment. All specifications include block fixed effects (corresponding to the stratification used in randomization) and school-level baseline controls. Robust standard errors clustered at the school level are reported in parentheses. The sample size varies across columns due to missing responses on specific study-time modules. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

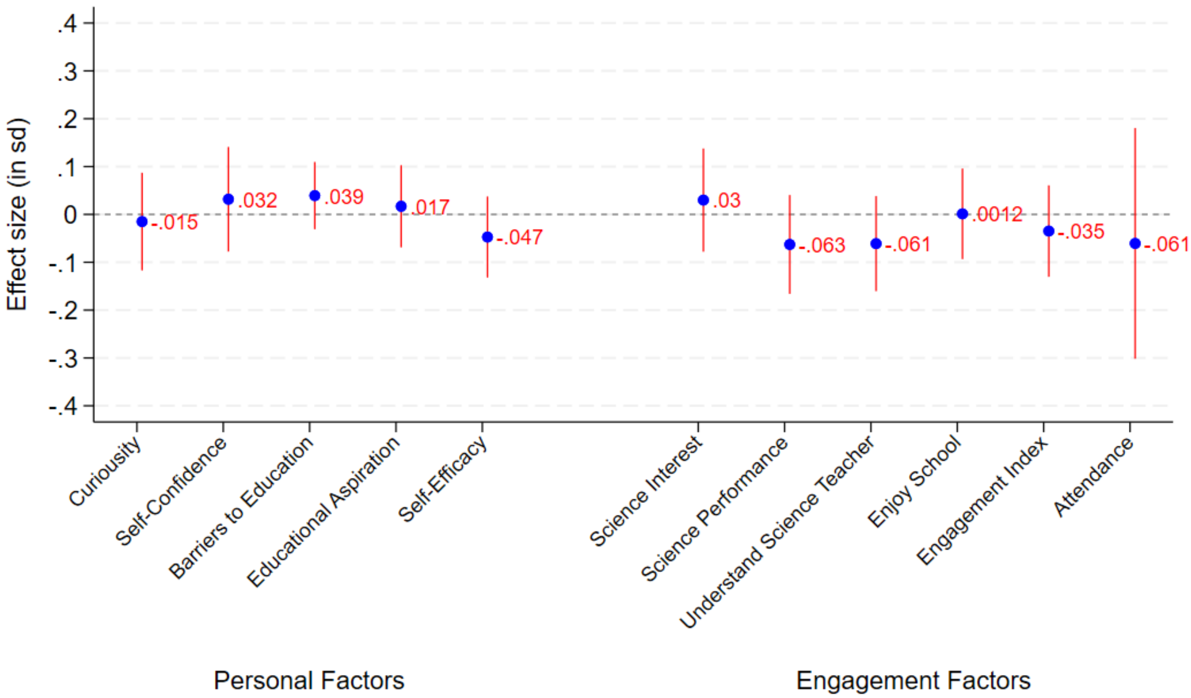
B Robustness Figures and Tables

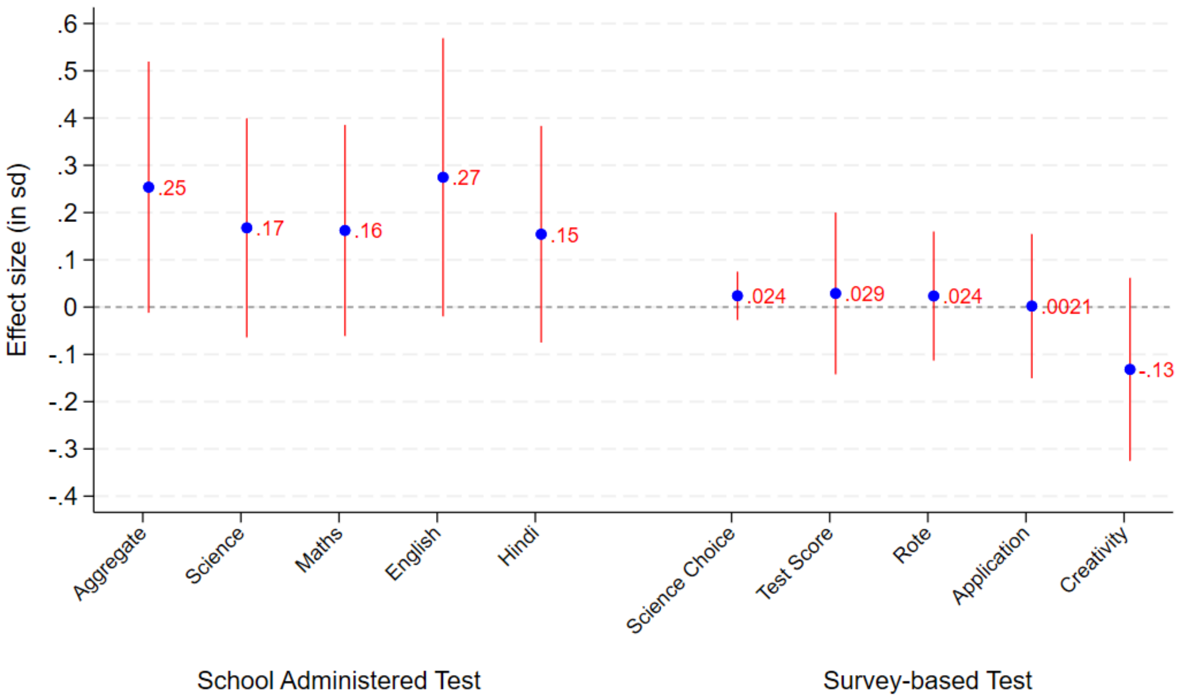
Figure B.1: Impact of Treatment (MSL): Controlling for student's characteristics



Notes: This figure presents the coefficient estimates after additionally adding student-level controls (gender, age, father's year of education, mother's year of education, household size, household wealth index, social desirability bias index). The explanatory variable is a binary indicator for MSL treatment assignment. All specifications include block fixed effects and school-level controls for baseline characteristics. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Figure B.2: Impact of Treatment (LoB): Controlling for student's characteristics





Notes: This figure presents the coefficient estimates after additionally adding student-level controls (gender, age, father's year of education, mother's year of education, household size, household wealth index, social desirability bias index). The explanatory variable is a binary indicator for MSL treatment assignment. All specifications include block fixed effects and school-level controls for baseline characteristics. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table B.1: Impact of the Mobile Science Lab on Personal Factors: Robustness to Social Desirability Bias

	Curiosity	Self-confidence	Barriers to education	Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
MSL	0.196**	0.171**	-0.241***	0.158***	0.090*
	(0.077)	(0.081)	(0.069)	(0.050)	(0.052)
SDB	0.133***	0.109***	-0.080***	0.102***	0.075***
	(0.020)	(0.020)	(0.020)	(0.011)	(0.015)
Control mean	0.000	0.000	0.000	0.000	0.000
Observations	5,750	5,750	5,750	5,386	4,994
R-squared	0.050	0.041	0.035	0.050	0.017
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Social desirability control	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55

Notes: This table replicates the baseline personal factors analysis controlling for social desirability bias (SDB), measured using the short-form Marlowe–Crowne Social Desirability Scale (see Appendix Section E.4). The sample includes 5,750 students across 55 schools (27 treatment, 28 control) in Uttar Pradesh. All dependent variables are normalized indices with mean zero and standard deviation one in the control group. The explanatory variable is a binary indicator for MSL treatment assignment. All specifications include block fixed effects and school-level controls for baseline characteristics. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table B.2: Impact of the Mobile Science Lab on Engagement Factors: Robustness to Social Desirability Bias

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
MSL	0.199***	0.161*	0.173**	0.199***	0.189**	-0.102
	(0.061)	(0.090)	(0.079)	(0.066)	(0.080)	(0.119)
SDB	0.140***	0.118***	0.143***	0.121***	0.143***	0.005
	(0.018)	(0.020)	(0.018)	(0.018)	(0.020)	(0.035)
Control mean	0.000	0.000	0.000	0.000	0.000	3.340
Observations	5,750	5,750	5,750	5,296	5,750	5,019
R-squared	0.054	0.046	0.059	0.041	0.053	0.021
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Social desirability control	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55	55

Notes: This table replicates the baseline student engagement analysis controlling for social desirability bias (SDB), measured using the short-form Marlowe–Crowne Social Desirability Scale (see Appendix Section E.4). The sample includes 5,750 students across 55 schools (27 treatment, 28 control) in Uttar Pradesh. All dependent variables are normalized indices with mean zero and standard deviation one in the control group, except Column (6) which reports attendance in levels. The explanatory variable is a binary indicator for MSL treatment assignment. All specifications include block fixed effects and school-level controls for baseline characteristics. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table B.3: Impact of the Mobile Science Lab on Non-Cognitive Measures: Robustness to Social Desirability Bias

	Friends	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness
	(1)	(2)	(3)	(4)	(5)	(6)
MSL	-0.116	-0.217	-0.105	-0.269**	-0.132	-0.176
	(0.145)	(0.156)	(0.128)	(0.118)	(0.118)	(0.126)
SDB	0.135***	0.046	-0.032	0.092**	0.006	0.008
	(0.038)	(0.033)	(0.027)	(0.036)	(0.028)	(0.026)
Control mean	4.350	4.360	4.490	4.650	4.740	4.470
Observations	5,145	5,166	5,167	5,169	5,168	5,169
R-squared	0.092	0.022	0.024	0.030	0.019	0.022
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Social desirability control	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	55	55	55	55	55	55

Notes: This table replicates the baseline non-cognitive measures analysis controlling for social desirability bias (SDB), measured using the short-form Marlowe–Crowne Social Desirability Scale (see Appendix Section E.4). The sample includes 5,750 students across 55 schools (27 treatment, 28 control) in Uttar Pradesh. All dependent variables are normalized indices with mean zero and standard deviation one in the control group. The explanatory variable is a binary indicator for MSL treatment assignment. Columns (2)–(6) report Big Five personality measures: extraversion, agreeableness, conscientiousness, emotional stability, and openness. All specifications include block fixed effects and school-level controls for baseline characteristics. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table B.4: Impact of the Lab on Bike on Personal Factors: Robustness to Social Desirability Bias

	Curiosity	Self-confidence	Barriers to education	Aspiration	Self-efficacy
	(1)	(2)	(3)	(4)	(5)
LoB	-0.002	0.040	0.034	0.030	-0.037
	(0.052)	(0.055)	(0.036)	(0.045)	(0.043)
SDB	0.184***	0.169***	-0.132***	0.120***	0.103***
	(0.015)	(0.016)	(0.015)	(0.015)	(0.017)
Control mean	0.000	0.000	0.000	0.000	0.000
Observations	8,453	8,453	8,453	7,708	7,069
R-squared	0.055	0.054	0.042	0.030	0.019
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Social desirability control	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76

Notes: This table replicates the baseline personal factors analysis for the Lab on Bike (LoB) sample, controlling for social desirability bias (SDB), measured using the short-form Marlowe–Crowne Social Desirability Scale (see Appendix Section E.4). The sample includes 8,453 students across 76 schools in Uttar Pradesh. All dependent variables are normalized indices with mean zero and standard deviation one in the control group. The explanatory variable is a binary indicator for LoB treatment assignment. All specifications include block fixed effects and school-level controls for baseline characteristics. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table B.5: Impact of the Lab on Bike on Engagement Factors: Robustness to Social Desirability Bias

	Science interest	Performance in science	Understand science teacher	Enjoy school	Engagement index	Attendance
	(1)	(2)	(3)	(4)	(5)	(6)
LoB	0.037	-0.053	-0.051	0.010	-0.026	-0.066
	(0.054)	(0.053)	(0.050)	(0.047)	(0.048)	(0.122)
SDB	0.175***	0.179***	0.179***	0.167***	0.215***	0.080***
	(0.014)	(0.017)	(0.016)	(0.012)	(0.014)	(0.026)
Control mean	0.000	0.000	0.000	0.000	0.000	3.310
Observations	8,453	8,453	8,453	7,433	8,453	7,331
R-squared	0.049	0.062	0.056	0.051	0.077	0.029
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Social desirability control	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76	76

Notes: This table replicates the baseline engagement analysis for the Lab on Bike (LoB) sample, controlling for social desirability bias (SDB), measured using the short-form Marlowe–Crowne Social Desirability Scale (see Appendix Section E.4). The sample includes 8,453 students across 76 schools in Uttar Pradesh. All dependent variables are normalized indices with mean zero and standard deviation one in the control group, except Column (6) which reports attendance in levels. The explanatory variable is a binary indicator for LoB treatment assignment. All specifications include block fixed effects and school-level controls for baseline characteristics. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table B.6: Impact of the Lab on Bike on Non-Cognitive Measures: Robustness to Social Desirability Bias

	Friends	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness
	(1)	(2)	(3)	(4)	(5)	(6)
LoB	0.009	-0.121	-0.195**	-0.126	-0.139	-0.167
	(0.155)	(0.125)	(0.086)	(0.122)	(0.124)	(0.103)
SDB	0.158***	0.054*	-0.020	0.038	-0.033	0.004
	(0.034)	(0.031)	(0.023)	(0.023)	(0.032)	(0.027)
Control mean	4.070	4.680	4.700	4.780	4.810	4.580
Observations	7,521	7,561	7,560	7,557	7,558	7,559
R-squared	0.046	0.026	0.033	0.039	0.048	0.039
Block fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Social desirability control	Yes	Yes	Yes	Yes	Yes	Yes
Number of clusters	76	76	76	76	76	76

Notes: This table replicates the baseline non-cognitive measures analysis for the Lab on Bike (LoB) sample, controlling for social desirability bias (SDB), measured using the short-form Marlowe–Crowne Social Desirability Scale (see Appendix Section E.4). The sample includes 8,453 students across 76 schools in Uttar Pradesh. All dependent variables are normalized indices with mean zero and standard deviation one in the control group. The explanatory variable is a binary indicator for LoB treatment assignment. Columns (2)–(6) report Big Five personality measures: extraversion, agreeableness, conscientiousness, emotional stability, and openness. All specifications include block fixed effects and school-level controls for baseline characteristics. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table B.7: Pre-Treatment Balance in School Infrastructure: Mobile Science Lab (MSL), 2020

Variable	Control Schools N	Control Schools Mean (SE)	MSL Schools N	MSL Schools Mean (SE)	Diff. (MSL – Control)
	(1)	(2)	(3)	(4)	(5)
Total enrollment (Grades 6–8)	28	131.68 (11.89)	30	147.93 (8.70)	16.26
Female enrollment share	28	0.51 (0.01)	30	0.49 (0.01)	-0.01
Scheduled Caste enrollment share	28	0.29 (0.03)	30	0.26 (0.03)	-0.03
Other Backward Class enrollment share	28	0.58 (0.04)	30	0.57 (0.04)	0.01
Total teachers	28	7.71 (0.76)	30	7.50 (0.73)	-0.21
Student–teacher ratio (Grades 6–8)	28	25.88 (4.60)	30	26.01 (3.01)	0.13
Female teachers share	28	0.41 (0.06)	30	0.42 (0.05)	0.01
Teachers with graduate degree or higher	28	6.89 (0.69)	30	7.13 (0.71)	0.24
Teachers with graduate degree or higher (per student)	28	0.06 (0.01)	30	0.05 (0.01)	-0.00
Non-teaching staff	28	0.61 (0.35)	30	0.60 (0.29)	-0.00
Non-teaching staff (per student)	28	0.01 (0.00)	30	0.00 (0.00)	0.00
Dummy for school boundary wall	28	0.54 (0.10)	30	0.77 (0.09)	0.23^{*}
Dummy for school boundary wall (concrete)	28	0.46 (0.10)	30	0.77 (0.08)	0.30^{**}
Number of buildings	28	2.54 (0.62)	30	2.63 (0.52)	-0.01
Number of buildings (concrete)	28	2.36 (0.60)	30	2.40 (0.52)	-0.04
Number of toilets for boys	28	1.46 (0.17)	30	1.40 (0.13)	-0.06
Number of toilets for girls	28	1.61 (0.17)	30	1.38 (0.10)	-0.24
Toilets (per student)	28	0.03 (0.00)	30	0.02 (0.00)	-0.01^{*}
Number of classrooms	28	7.75 (0.82)	30	8.00 (0.82)	0.25
Number of classrooms (in good condition)	28	5.82 (0.75)	30	5.73 (0.74)	-0.09
Dummy for school playground	28	0.75 (0.08)	30	0.73 (0.08)	-0.02
Dummy for school with electricity	28	0.93 (0.05)	30	0.80 (0.07)	-0.13
Dummy for medical checkup facility	28	0.36 (0.09)	30	0.30 (0.09)	-0.06
Free textbooks distributed (per student)	28	0.61 (0.10)	30	0.60 (0.08)	-0.00
Free uniforms distributed (per student)	28	0.48 (0.09)	30	0.60 (0.08)	0.13
School Infrastructure Index	28	0.00 (0.19)	30	-0.53 (0.23)	-0.53^{*}

Notes: This table reports pre-intervention balance in school-level characteristics for the Mobile Science Lab (MSL) intervention using administrative data from UDISE+ (2020), which were used for stratification and randomization. The analysis sample includes the full original sample of 30 treatment and 28 control schools prior to any attrition. All variables are measured at the school level prior to treatment. The characteristics reported include enrollment and composition measures (total enrollment in Grades 6–8, female share, and caste composition), teacher characteristics (total number of teachers, share of female teachers, number of teachers with a graduate degree or higher, and non-teaching staff), and school infrastructure indicators (boundary walls, buildings and classrooms, toilet facilities, playgrounds, electricity, medical checkup facilities, and distribution of free textbooks and uniforms per student). The school infrastructure index aggregates indicators for boundary wall, playground, electricity, medical checkup facility, and library presence. Reported values are means with standard errors in parentheses. Column (5) reports differences in means between treatment and control schools, along with two-sided *t*-tests of equality of means. ***, **, * mean statistical significance at 1, 5, 10 percent levels.

Table B.8: Pre-Treatment Balance in School-Level Characteristics: Lab on Bike (LoB), 2020

Variable	Control Schools		LoB Schools		Diff. (LoB – Control)
	N	Mean (SE)	N	Mean (SE)	
	(1)	(2)	(3)	(4)	(5)
Total enrollment (Grades 6–8)	36	140.31 (9.58)	42	144.24 (7.15)	3.94
Female enrollment share	36	0.49 (0.02)	42	0.49 (0.01)	0.00
Scheduled Caste enrollment share	36	0.36 (0.03)	42	0.41 (0.03)	0.05
Other Backward Class enrollment share	36	0.57 (0.03)	42	0.54 (0.03)	-0.03
Total teachers	36	7.58 (0.79)	42	9.60 (0.69)	2.01*
Student–teacher ratio (Grades 6–8)	36	22.50 (1.73)	42	18.40 (1.58)	-4.09*
Female teachers share	36	0.64 (0.05)	42	0.65 (0.04)	0.00
Teachers with graduate degree or higher	36	6.89 (0.75)	42	8.91 (0.65)	2.02**
Teachers with graduate degree or higher (per student)	36	0.05 (0.00)	42	0.07 (0.00)	0.16**
Non-teaching staff	36	1.17 (0.55)	42	2.33 (0.69)	1.17
Dummy for school boundary wall	36	0.83 (0.06)	42	0.81 (0.06)	-0.02
Dummy for school boundary wall (concrete)	36	0.83 (0.06)	42	0.79 (0.06)	-0.05
Number of buildings	36	2.58 (0.49)	42	3.10 (0.40)	0.54
Number of buildings (concrete)	36	2.50 (0.43)	42	3.02 (0.40)	0.52
Number of toilets for boys	36	1.58 (0.14)	42	2.01 (0.22)	0.51*
Number of toilets for girls	36	1.78 (0.16)	42	2.24 (0.20)	0.46*
Toilets (per student)	35	0.03 (0.00)	42	0.03 (0.00)	0.01
Number of classrooms	36	6.81 (0.78)	42	7.69 (0.74)	0.86
Number of classrooms (in good condition)	36	5.94 (0.79)	42	6.14 (0.78)	0.20
Dummy for school playground	36	0.72 (0.08)	42	0.74 (0.07)	0.02
Dummy for school with electricity	36	0.86 (0.06)	42	0.95 (0.03)	0.09
Dummy for medical checkup facility	36	0.47 (0.08)	42	0.57 (0.07)	0.10
Free textbooks distributed (per student)	36	0.84 (0.06)	42	0.87 (0.04)	0.03
Free uniforms distributed (per student)	36	0.79 (0.07)	42	0.86 (0.05)	0.06
School Infrastructure Index	36	0.00 (0.17)	42	0.18 (0.18)	-0.18

Notes: This table reports pre-intervention balance in school-level characteristics for the Lab on Bike (LoB) intervention using administrative data from UDISE+ (2020), which were used for stratification and randomization. The analysis sample includes the full original sample of 42 treatment and 36 control schools prior to any attrition. All variables are measured at the school level prior to treatment. Enrollment refers to total enrollment in Grades 6–8. Caste shares correspond to the proportion of enrolled students belonging to Scheduled Castes and Other Backward Classes. Teacher characteristics include the total number of teachers, the share of female teachers, and the number of teachers with a graduate degree or higher. The student–teacher ratio is computed using enrollment and teacher counts for Grades 6–8. Infrastructure variables capture the presence and quality of school facilities, including buildings, classrooms, toilets, playgrounds, electricity, and medical checkup facilities, as well as the distribution of free textbooks and uniforms per student. Reported values are means with standard errors in parentheses. Column (5) reports differences in means between treatment and control schools, along with two-sided t-tests of equality of means. ***, **, * mean statistical significance at 1, 5, 10 percent levels.

Table B.9: Are suspicious classes more in treatment schools?

	Suspicious class (MSL)	Suspicious class (LoB)
	(1)	(2)
Treatment (MSL/LoB)	0.052	0.136
	(0.046)	(0.086)
Control mean	0.07	0.19
Observations	165	212
R-squared	0.202	0.107
Block fixed effects	Yes	Yes
Controls	Yes	Yes
Number of clusters	55	72

Notes: This table reports the probability of a suspicious class (created in the Appendix D) falling in the treatment group - Mobile Science Lab (MSL) in Column (1) and Lab on Bike (LoB) in Column (2). The analysis is at the school-class level: 165 in the MSL area (81 treatment and 84 control) and 212 in the LoB area (118 treatment and 90 control). The dependent variable is a dummy for suspicious class- high average school-conducted test scores, low variance and low Herfandhl index. The explanatory variable is a binary indicator for schools assigned to receive the MSL (or LoB) treatment. All specifications include block fixed effects (corresponding to the stratification used in randomization), and school-level controls for baseline characteristics. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table B.10: Impact of Mobile Science Lab on Administrative Test Scores: Excluding suspicious class

	Aggregate	Science	Maths	English	Hindi
	(1)	(2)	(3)	(4)	(5)
MSL	0.361**	0.269**	0.329**	0.304**	0.210
	(0.147)	(0.124)	(0.124)	(0.127)	(0.141)
Control mean	0.000	0.000	0.000	0.000	0.000
Observations	7,133	7,133	7,133	7,133	7,133
R-squared	0.196	0.159	0.137	0.155	0.172
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Grade fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	54	54	54	54	54

Notes: This table reports the impact of the Mobile Science Lab (MSL) treatment on students' performance in school-administered examinations using administrative test score data. The analysis sample excludes the classes with potential manipulation during exam (suspicious cluster of classes), hence the sample drops from 7,899 students across 55 schools. All dependent variables are normalized test scores with mean zero and standard deviation one in the control group, derived from school-conducted examinations in four core subjects. The explanatory variable is a binary indicator for schools assigned to receive the MSL treatment. All specifications include block fixed effects (corresponding to the stratification used in randomization), grade fixed effects to account for differences across Grades 6–8, and school-level controls for baseline characteristics. Column (1) reports treatment effects on the composite aggregate score index, which combines performance across all four subjects. Column (2) reports effects on the science test score, the primary subject targeted by the MSL intervention. Column (3) to Column (5) report effects on the mathematics, English and Hindi test score, respectively. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at 1, 5, and 10 percent levels.

Table B.11: Impact of the Lab on Bike on Administrative Test Scores: Excluding suspicious class

	Total	Science	Maths	English	Hindi
	(1)	(2)	(3)	(4)	(5)
LoB	0.113	0.030	0.076	0.198	0.049
	(0.130)	(0.115)	(0.120)	(0.149)	(0.107)
Control mean	0.000	0.000	0.000	0.000	0.000
Observations	8,081	8,081	8,081	8,081	8,081
R-squared	0.364	0.316	0.239	0.213	0.338
Block fixed effects	Yes	Yes	Yes	Yes	Yes
Grade fixed effects	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes
Number of clusters	68	68	68	68	68

Notes: This table reports the impact of the Lab on Bike (LoB) treatment on students' performance in school-administered examinations using administrative test score data. The analysis sample excludes the classes with potential manipulation during exam (suspicious cluster of classes), hence the sample drops from 10,289 students across 72 schools. All dependent variables are normalized to have mean zero and standard deviation one in the control group. The explanatory variable is a binary indicator for schools assigned to receive the LoB treatment. All specifications include block fixed effects (corresponding to the stratification used in randomization), grade fixed effects to account for differences across Grades 6–8, and school-level controls for baseline characteristics. Column (1) reports treatment effects on the composite aggregate score index, which combines performance across all four subjects. Column (2) reports effects on the science test score, the primary subject targeted by the MSL intervention. Column (3) to Column (5) report effects on the mathematics, English and Hindi test score, respectively. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table B.12: Are Missing Values Correlated with the Treatment Indicator (MSL)?

	Aspiration	Self-Efficacy	Creativity	Friends	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
MSL	0.004	0.036	0.012	0.013	0.015	0.014	0.015	0.014	0.015
	(0.009)	(0.025)	(0.022)	(0.025)	(0.025)	(0.026)	(0.026)	(0.026)	(0.026)
Control Mean	0.07	0.12	0.07	0.10	0.09	0.09	0.09	0.09	0.09
Observations	5,750	5,750	5,750	5,750	5,750	5,750	5,750	5,750	5,750
R-squared	0.007	0.012	0.036	0.056	0.044	0.044	0.044	0.044	0.044
Block Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Clusters	55	55	55	55	55	55	55	55	55

Notes: This table reports whether missing values in responses are correlated with treatment assignment in the Mobile Science Lab (MSL) sample. The analysis sample includes 5,750 students across 55 schools (27 treatment and 28 control schools) in Uttar Pradesh. All dependent variables are binary indicators equal to one if the outcome is missing, either because the student marked more than one bubble or left the item blank. The explanatory variable is a binary indicator for schools assigned to receive the MSL treatment. All specifications include block fixed effects corresponding to the randomization strata and school-level controls for baseline characteristics. Columns (1)–(4) report results for educational aspiration, self-efficacy, creativity, and number of friends, respectively. Columns (5)–(9) report results for the Big Five personality traits: extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience. Control means indicate the share of students for whom the relevant outcome is missing. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

Table B.13: Are Missing Values Correlated with the Treatment Indicator (LoB)?

	Aspiration	Self-Efficacy	Creativity	Friends	Extraversion	Agreeableness	Conscientiousness	Emotional Stability	Openness
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
LoB	−0.038*	−0.004	−0.018	0.033	0.023	0.023	0.023	0.022	0.022
	(0.020)	(0.027)	(0.017)	(0.025)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Control Mean	0.10	0.16	0.10	0.09	0.09	0.09	0.09	0.09	0.09
Observations	8,453	8,453	8,453	8,453	8,453	8,453	8,453	8,453	8,453
R-squared	0.045	0.040	0.058	0.069	0.074	0.074	0.075	0.076	0.075
Block Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Clusters	76	76	76	76	76	76	76	76	76

Notes: This table reports whether missing values in responses are correlated with treatment assignment in the Lab on Bike (LoB) sample. The analysis sample includes 8,453 students across 76 schools (41 treatment and 35 control schools) in Uttar Pradesh. All dependent variables are binary indicators equal to one if the outcome is missing, either because the student marked more than one bubble or left the item blank. The explanatory variable is a binary indicator for schools assigned to receive the LoB treatment. All specifications include block fixed effects corresponding to the randomization strata and school-level controls for baseline characteristics. Columns (1)–(4) report results for educational aspiration, self-efficacy, creativity, and number of friends, respectively. Columns (5)–(9) report results for the Big Five personality traits: extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience. Control means indicate the share of students for whom the relevant outcome is missing. Robust standard errors clustered at the school level are reported in parentheses. ***, **, * denote statistical significance at the 1, 5, and 10 percent levels.

C Pedagogical Design

The learning arc. Sessions open with a *super start*—a five- to seven-minute activity designed to provoke surprise before any explanation is offered. In a session on light, for example, students receive periscopes to explore prior to any discussion of reflection or angles.[25] The questions students generate are written on the blackboard and become the organizing thread of the lesson. This inversion—phenomenon before principle, observation before abstraction—is the structural mechanism through which the curriculum activates curiosity and establishes inquiry as the driver of learning (Alan and Mumcu, 2024; Nourani et al., 2025).

Experimentation and reasoning. Students work in groups of four to six. They predict outcomes before each experiment and articulate observations in their own words using action-oriented scientific verbs—*observe, compare, measure, classify, infer*. The predict-then-verify structure simultaneously deepens content understanding through the direct cognitive pathway and builds perceived competence and self-efficacy through the indirect motivational pathway (Bando et al., 2019; Fredricks et al., 2004). Every student conducts each experiment independently; the experience of completing a task successfully is intended to reinforce both mastery beliefs and persistence (Alan et al., 2019).

Synthesis and extension. Sessions conclude with a *super finish* in which the instructor uses concise "key message" charts to connect empirical observations to formal terminology, consolidating the session's conceptual content. Learning is extended through short "Do It Yourself" prompts encouraging home replication, accompanied by a brief story about a scientific discovery or scientist that illustrates creativity and perseverance. Each session typically includes one such story integrated into the logical flow of the topic.

Curricular fit and monitoring. Topics are sequenced to reinforce material taught in regular classes, so experimental sessions serve as complements rather than substitutes for standard instruction, occupying a small fraction of total instructional time while serving as conceptual anchors. Instructors follow a standardized pedagogical script—*super start, inquiry, experimentation, synthesis*—while adapting examples and pacing to local context; formal lecturing and extended note-taking are avoided. These design choices reflect the autonomy-support lever: by minimizing didactic instruction and providing structured opportunities for peer dialogue and self-directed exploration, the curriculum reinforces students' sense of agency over learning (Reeve, 2012; Skinner and Pitzer, 2012). Agastya's internal monitoring system conducts random supervisory visits to verify adherence.

Dosage and learning dynamics. Because the intervention embeds substantive science content within experiential sessions, motivational and engagement shifts are likely to emerge gradually through repeated exposure rather than from any single session. Content learning may accumulate more directly with attendance, while shifts in competence beliefs and intrinsic motivation may require threshold levels of exposure before

[25] Periscopes are uncommon in rural India and therefore serve as effective curiosity triggers.

they materialize—implying that learning gains need not scale linearly with dosage. These dynamics are relevant for interpreting differences in effectiveness between the two delivery arms.

MSL versus LoB. Both arms deliver the same curriculum through the same pedagogical sequence. MSL operates from a mobile van with a team of instructors and achieves higher visit frequency; LoB delivers identical content through a single instructor traveling by motorbike. These logistical differences generate variation in cumulative exposure—the key margin of treatment intensity—without altering the underlying pedagogy (Table 1).

D Suspicious Classrooms with School-Based Test Scores

We first calculate three class-level statistics (mean, variance, and homogeneity index), followed by principal component analysis to extract the first two principal components. These are then used as inputs into k-means clustering to generate six clusters with varying levels of the three statistics. Unlike Singh (2024), we do not have question-level responses, implying that we cannot distinguish between unattempted and incorrectly answered questions. However, we have subject-wise number of questions and subject-wise scores. Since there was a single question paper covering all subjects, we treat each subject as one item to compute the homogeneity index.

In the MSL group, the total number of questions in Mathematics, English, and Hindi is four each; in Science it is five (for grades 6 and 7) or six (grade 8). Correspondingly, the possible set of obtained marks is $(0, 1, 2, 3, 4)$ in Mathematics, English, and Hindi, and $(0, 1, 2, 3, 4, 5, 6)$ in Science. We compute the class-level homogeneity index as a weighted average of the Gini index for each subject:

$$H_i = \sum_{s \in \{M,E,H,Sc\}} w_s H_{is}$$

where w_s is the weight based on the number of questions for subject s. For example, w_M (the Mathematics weight) is 0.235 ($= 4/(4+4+4+5)$) for grades 6 and 7, and 0.222 ($= 4/(4+4+4+6)$) for grade 8. H_{is} is the Gini index for subject s in class i, computed as:

$$H_{is} = 1 - \sum_{m=0}^{4/5/6} \left(\frac{n_{m,s}}{N_i}\right)^2$$

where $\frac{n_{m,s}}{N_i}$ is the share of students in class i obtaining marks m in subject s. The index equals zero if all students in class i receive the same marks in subject s, implying no heterogeneity. From these statistics, we extract the first two principal components and use them to create six clusters. To avoid convergence to a local minimum, we run the clustering algorithm 100 times and retain the solution with the lowest within-cluster sum of squares. The cluster-wise average statistics are reported below.

Based on the above, we assign classes falling under cluster 5 in the MSL and cluster 1 in the LoB as

Table B.14: Cluster-wise averages in MSL and LoB

MSL					LoB				
Cluster	Mean (Score)	Variance (Score)	Homogeneity Index	Share Class	Cluster	Mean (Score)	Variance (Score)	Homogeneity Index	Share Class
1	67.12	225.59	0.64	0.21	**1**	**82.54**	**37.11**	**0.19**	**0.24**
2	76.81	129.94	0.48	0.19	2	74.73	113.55	0.42	0.28
3	18.40	105.03	0.45	0.07	3	3.28	23.92	0.18	0.06
4	47.15	206.25	0.67	0.25	4	58.96	806.72	0.67	0.06
5	**81.11**	**67.34**	**0.29**	**0.10**	5	70.24	260.8	0.62	0.18
6	54.99	513.19	0.72	0.19	6	53.66	162.11	0.6	0.17
Total	*59.87*	*232.91*	*0.58*	*165*	*Total*	*66.76*	*167.74*	*0.43*	*212*

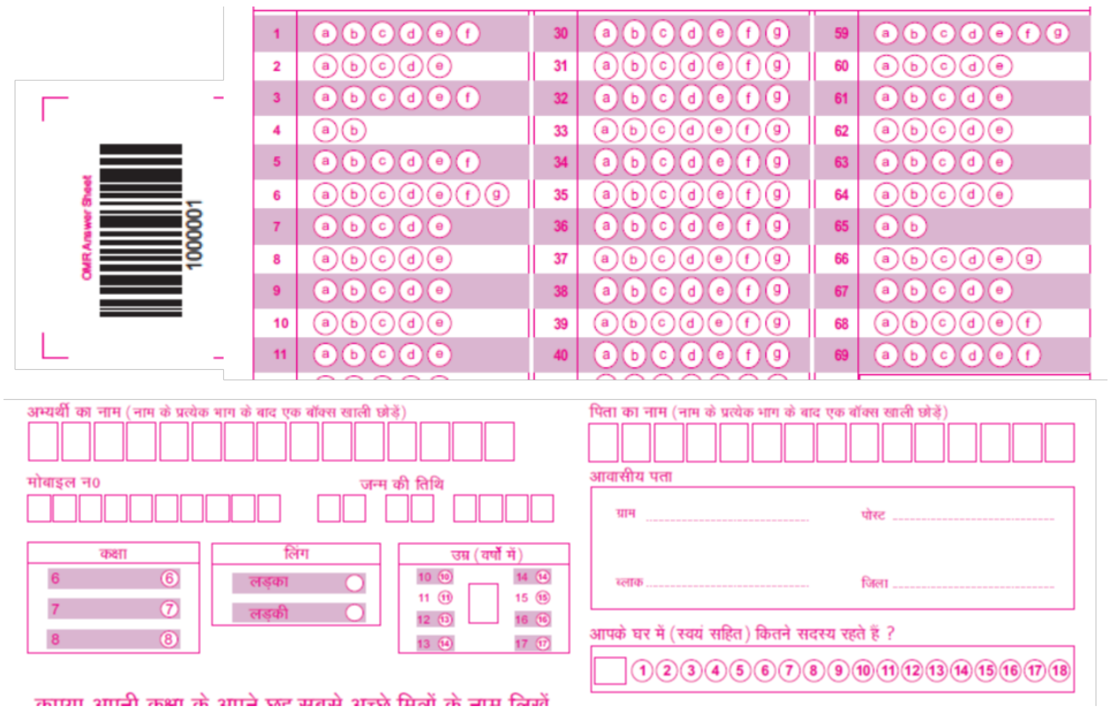
Notes: This table reports the averages of the three statistics (mean, variance, and Gini index), and the share of classes in each cluster for MSL and LoB separately. The bold entries indicate the suspicious clusters (cluster 5 in MSL and cluster 1 in LoB), characterized by high average scores, low variance, and a low homogeneity index.

suspicious. These clusters exhibit very high average scores, low variance, and very low homogeneity in students' marks. They account for 10% of classes in the MSL arm and 24% in the LoB arm.

E Survey Instruments and Measures

E.1 Survey Response Format

Figure F.1: Sample Optical Mark Recognition (OMR) Response Sheet



Notes: This figure shows an example of the Optical Mark Recognition (OMR) response sheet used to record student answers in the endline survey. Students marked their responses directly on the sheet, which was subsequently scanned and digitized for analysis.

E.2 Student Survey

The student survey was administered to students in Grades 6–8. The survey consisted of multiple questions designed to capture students' educational inputs, attitudes toward science, non-cognitive skills, engagement, and household background. All questions were presented in multiple-choice format to ensure compatibility with OMR-based data collection. Below we describe the main modules of the survey.

Module A : Schooling History, Attendance, and Subject Preferences

This module collected information on grade repetition, recent school absenteeism, students' favorite subjects, availability of homework support at home, and engagement with science-related activities outside school, such as watching science programs or conducting experiments at home.

Q1. Last year you were in which class?

a) Class 5

b) Class 6

c) Class 7

d) Class 8

Q2. Have you ever repeated a grade in school?

a) Yes

b) No

c) Don't want to answer

Q3. Past week, how often were you absent from school?

a) Never

b) 1 day

c) 2 days

d) 3 days

e) 4 days

f) 5 days or full week

Q4. Have you ever done a science-based experiment?

a) Yes

b) No

c) I do not know what a science-based experiment is

Q5. What are your two most favorite subjects? *[Select 1 or 2]*

a) Hindi

b) Science

c) English

d) Maths

e) Social Science (history, geography, civics, etc.)

Q6. Is there someone at home who teaches you or helps you in doing homework? *[Multiple select allowed]*

a) Father

b) Mother

c) Elder brother

d) Elder sister

e) Others (uncle, aunt, grandparents, etc.)

f) No one helps

Q7. In my leisure time, I like . . . *[Multiple select allowed]*

a) Watching popular science channels on TV

b) Watching other science-related videos on television or through the internet

c) Reading science-related books or magazines (other than school book)

d) Conducting homemade science experiments

e) Having conversations with others about science

f) None of the above

Module B : Time Use and Study Inputs

Students reported the amount of time spent on household work, self-study, and private tuition. Subject-specific time use was elicited for science, mathematics, and other subjects, distinguishing between tuition-based study and independent study at home.

How many hours, on an average weekday, do you spend doing the following (in one day)?

Q8. Help family with household work or other types of work (e.g., agriculture, cooking, cleaning, caring for younger siblings or elders)

a) Almost no time in household work

b) Less than 1 hour

c) 1 to 2 hours

d) 2 to 3 hours

e) More than 3 hours

Q9. Study (do homework) outside of tuition and school per day?

a) Don't study at home
b) Less than 1 hour
c) 1 to 2 hours
d) 2 to 3 hours
e) More than 3 hours

Q10. Time spent per day on tuition (please do not add travel time)

a) Don't go for tuition *(if this, skip to Q18)*
b) Less than 1 hour
c) 1 to 2 hours
d) 2 to 3 hours
e) More than 3 hours

Last week, how many hours per day did you spend in private tuition for the following subjects?

Q11. Science **Q12.** Maths **Q13.** Other subjects

a) 0–2 hours
b) 2–4 hours
c) 4–6 hours
d) 6–8 hours
e) More than 8 hours

Last week, how many hours did you spend studying at home (do not include tuition time)?

Q14. Science **Q15.** Maths **Q16.** Other subjects

a) 0–2 hours
b) 2–4 hours
c) 4–6 hours
d) 6–8 hours
e) More than 8 hours

Module C : Educational Aspirations and Perceived Constraints

This module measured students' educational aspirations, perceived likelihood of completing higher education, intended subject streams, and perceived barriers to educational attainment, including financial constraints, family expectations, social identity, and self-confidence.

Questions related to what you think about yourself. Please select the best possible option. There is no right or wrong answer.

Q17. Thinking about your future education-wise, what objective do you intend to achieve?

a) Up to Class 8

b) Up to Class 10

c) Up to Class 12

d) Up to a university degree

Q18. Independently from your educational aim, but thinking about your abilities, do you think you could obtain a university degree?

a) Not at all

b) Very little

c) Somewhat

d) Very much

Q19. Thinking about your future education-wise, what subject do you intend to study in higher grades?

a) Science stream (science, maths-type courses)

b) Arts stream (English, history, geography-type courses)

c) Commerce stream (business, bookkeeping-type courses)

d) I have not made up my mind yet

e) Don't know

Do you think the following could be an obstacle in the achievement of your educational aim?

Q20. Economic resources **Q21.** Family needs and ideas **Q22.** Gender/caste prejudice **Q23.** Not feeling up to the standards

a) Very much

b) Somewhat

c) Very little

d) Not at all

Module D : Attitudes Toward Science Learning

Students rated their enjoyment of learning science, perceived relevance of science, and interest in conducting experiments using a 7-point Likert scale (1 = Strongly Disagree to 7 = Strongly Agree).

How much do you agree with these statements about learning science?

Q24. I enjoy learning science.
Q25. Science teaches me how things in the world work.
Q26. I like to do science experiments.

Response scale (all items in this module):

a) Strongly Disagree

b) Disagree

c) Somewhat Disagree

d) Neither Agree Nor Disagree

e) Somewhat Agree

f) Agree

g) Strongly Agree

Module E : Perceptions of Science Instruction

This module captured students' perceptions of teaching quality, including clarity of instruction, responsiveness to questions, use of varied teaching methods, and willingness to re-explain concepts.

How much do you agree with these statements about your science lessons? (Same 7-point scale as Module D.)

Q27. My teacher is easy to understand.
Q28. My teacher has clear answers to my questions.
Q29. My teacher is good at explaining science.
Q30. My teacher does a variety of things (e.g., conducting small experiments, providing examples, etc.) to help us learn.
Q31. My teacher explains a topic again when we don't understand.

Module F : Academic Self-Concept in Science

Students reported their perceived ability and confidence in science, including beliefs about doing well in science, learning speed, perceived difficulty, and comfort asking questions.

How much do you agree with these statements about science? (Same 7-point scale as Module D.)

Q32. I usually do well in science.

Q33. I learn things quickly in science.

Q34. Science is harder for me than any other subject.

Q35. Science is not confusing for me.

Q36. I do not feel shy asking questions during my science class. *(Note: response scale is reversed in the original instrument relative to Q32–Q35; higher agreement = more confidence.)*

Module G : Household Background and Assets

This module collected information on parental education and occupation, sibling composition, housing characteristics, and ownership of household assets.

Q37. Father's primary occupation

a) Farming

b) Non-farm business (cloth shop, selling milk, etc.)

c) Self-employment (blacksmith, barber, etc.)

d) Government or private job

e) Unemployed/Retired

f) Other

Q38. Father's educational level **Q39.** Mother's educational level

a) No formal schooling

b) Up to Class 5

c) Up to Class 8

d) Up to Class 10

e) Up to Class 12

f) College or above

g) Don't know

Q40. Do you have any brother(s)/sister(s) studying in the same school as you?

a) Yes

b) No

Q41. In which class are your brother(s)/sister(s) studying? *[Multiple select allowed]*

a) Class 5 or less

b) Class 6

c) Class 7

d) Class 8

e) Above Class 8

f) Not applicable

Tell us about the house you are currently living in.

Q42. Number of rooms (having 4 walls; not including kitchen/bathroom/common living area)

a) 1

b) 2–3

c) 4–5

d) 6–7

e) More than 7

Q43. Main material of the roof

a) Pucca (concrete)

b) Semi-pucca (khappad/tin)

c) Kutcha

How many of the following items are there in your household?

Q44. Motor vehicles *[Multiple select allowed]*

a) Car/Jeep

b) Tractor

c) Motorcycle

d) Cycle

e) Nothing

Q45. Phone *[Multiple select allowed]*

a) Smartphone

b) Simple phone/landline

c) Nothing

d) Don't know

Q46. Television *[Multiple select allowed]*

a) Color TV

b) Black-and-white TV

c) No TV

Q47. Number of cattle (cow/buffalo)

a) 0

b) 1–2

c) 3–4

d) 5–6

e) More than 6

Module H : Mindset, Curiosity, and Non-Cognitive Skills

Standardized batteries were used to measure growth mindset, curiosity, self-confidence, and personality traits. These included fixed versus growth mindset items, anxiety-related items, a short Big Five personality battery, and components of a critical-thinking disposition scale.

Growth Mindset (Q48–Q50). *How much do you agree/disagree with the following? (5-point scale: Strongly Disagree to Strongly Agree)*

Q48. You have a certain amount of intelligence, and you can't really do much to change it.
Q49. Your intelligence is something about you that you can't change very much.
Q50. You can learn new things, but you can't really change your basic intelligence.

Anxiety Items (Q51–Q52). **Q51.** Even a small doubt can stop me from doing new things.
Q52. When I am unsure about anything, I get tensed.

Birth Order (Q53). **Q53.** Are you the first-born child to your parents?

a) Yes

b) No

Big Five Personality—10-Item Scale (Q54–Q63). *I see myself as . . . (7-point scale: Strongly Disagree to Strongly Agree)*

Q54. Extraverted, enthusiastic.
Q55. Critical, quarrelsome.
Q56. Dependable, self-disciplined.
Q57. Anxious, easily upset.
Q58. Open to new experiences, complex.
Q59. Reserved, quiet.
Q60. Sympathetic, warm.
Q61. Disorganized, careless.
Q62. Calm, emotionally stable.
Q63. Conventional, uncreative.

Curiosity—Critical-Thinking Disposition Scale (Q64–Q68). *(7-point scale: Strongly Disagree to Strongly Agree)*

Q64. I am curious because I try to find out what I don't know.
Q65. I like to learn how everything works.
Q66. I enjoy learning new things.
Q67. I don't care if I know more or less things. *(reverse-coded)*
Q68. I try to understand things as well as possible.

Self-Confidence—Critical-Thinking Disposition Scale (Q69–Q74). *(7-point scale: Strongly Disagree to Strongly Agree)*

Q69. I try to say what I think even if others think otherwise.
Q70. I am able to understand things by myself.
Q71. I trust what others think more than what I think. *(reverse-coded)*
Q72. I notice that I reason better and better.
Q73. I like learning something without the help of others.
Q74. Learning to think well is important in order to improve.

Module I : Program Exposure and Experience (Treatment Schools Only)

Students in treatment schools reported whether they attended Agastya-based sessions, the number of sessions attended, overall experience ratings, and which aspects they found most enjoyable or novel.

Q75. Did you attend any Agastya-based sessions?

a) Yes

b) No

Q76. How many sessions did you attend?

a) 1

b) 2

c) 3

d) 4

e) 5

f) 6

g) 7

Q77. Rate your experience during those sessions on average. *(5-point scale)*

a) Very good

b) Good

c) Okay, like any other class

d) Bad

e) Very bad

Q78. Which aspect(s) of the sessions did you enjoy the most?

a) Group activity

b) Science experiments

c) Real-life examples

d) Speaking up in the classroom

e) Teachers giving more time to speak in the classroom

f) Didn't enjoy

Q79. Do you want more such sessions?

a) Yes

b) No

Q80. Have you discussed these sessions with your parents, friends, or siblings?

a) Yes

b) No

Module J : Student Engagement and Effort

This module measured affective, behavioral, and cognitive engagement in school, including enjoyment of school, effort in class, participation, attention, homework completion, and help-seeking behavior.

How much do you agree with the following statements? (Same 7-point scale as Module D.)

Q81. I enjoy going to school.
Q82. I work as hard as I can in class.
Q83. I try very hard to do well in school.
Q84. I participate in class discussions.
Q85. I pay attention in class.
Q86. I complete my homework on time.
Q87. I ask questions when I don't understand something in my readings or during lectures.
Q88. I study at home even when I don't have a test.

Module K : Curriculum-Aligned Science Assessment (Grade-Specific)

The endline survey also included curriculum-aligned multiple-choice science questions separately for Grades 6, 7, and 8. Each grade-level assessment consisted of ten questions covering core concepts from the state syllabus, including physical sciences, life sciences, and environmental science.

Response Scales and Index Construction

Most survey items were recorded using ordered categorical response scales. Unless otherwise noted, higher values correspond to higher levels of the underlying construct.

Likert-Scale Items (Attitudes, Beliefs, and Engagement). Items measuring attitudes toward science, perceptions of instruction, academic self-concept, and student engagement were collected using Likert-type response scales ranging from four-point to seven-point (e.g., *strongly disagree* to *strongly agree*). For each construct, responses were coded numerically, standardized using the control-group mean and standard deviation, and averaged to form composite indices.

Non-Cognitive Skills. Growth mindset, curiosity, self-confidence, and personality traits were measured using short batteries adapted from established instruments. Growth mindset and curiosity items used agreement-based Likert scales, while personality traits were elicited using the 10-item Big Five battery. Individual items were standardized and aggregated into domain-specific indices following standard practice.

Educational Aspirations and Perceived Constraints. Items capturing aspirations and perceived barriers to attainment were recorded on ordered categorical scales reflecting likelihood or agreement and analyzed both individually and as part of standardized indices.

Time Use and Study Inputs. Time-use measures were collected in categorical bins reflecting hours spent on study, tuition, and household work. These variables are used in their original categorical form or converted to approximate continuous measures using bin midpoints.

Program Exposure Measures. Program exposure and experience questions were asked only in treatment schools and recorded using categorical or ordered response options. These measures characterize participation intensity and engagement with the intervention rather than serving as identifying variation.

Curriculum-Aligned Science Questions. Grade-specific science questions were scored as binary indicators of correct responses and aggregated to form standardized test-score indices within each grade.

E.3 Teacher Survey

The teacher survey was administered to science teachers in both treatment and control schools to document baseline instructional environments, teaching practices, and pedagogical beliefs, and to examine whether the intervention influenced reported teaching behaviors. The survey was administered using a structured questionnaire with ordered categorical response options. Below we summarize the main modules and response scales. Question numbering refers to the original survey instrument.

Module A : Demographic and Educational Background

This module collected information on teachers' gender, age, highest level of education completed, field of study, and social category. These variables are used to describe the teacher sample and to assess balance across treatment and control schools.

Q1. Are you female or male?

a) Female

b) Male

Q2. How old are you? *[Write a number]* Years: __________

Q3. What is the highest level of formal education you have completed?

a) Below Graduate degree

b) Graduate

c) Post Graduate

d) Bachelors in Education (or equivalent degree)

e) Masters in Education (or equivalent degree)

f) Other professional courses for teaching (Diploma, etc.)

Q4. For the highest degree, what was your major?

a) Science

b) Arts

c) Business and Economics

d) Engineering and Computer Science

e) Education

f) Other

Q5. Which social category do you belong to?

a) General Class

b) Scheduled Caste/Scheduled Tribe

c) Other Backward Caste

d) Do not want to respond

Module B : Teaching Assignment and Experience

Teachers reported the grades and subjects they currently teach, years of experience as a science teacher, and whether teaching was their first career choice. These measures capture exposure to science instruction and accumulated teaching experience.

Q6. In the current school, which grades do you teach science lessons to? *[Multiple select allowed]*

a) Grade 6

b) Grade 7

c) Grade 8

d) Other than Grade 6, 7 and 8

Q7. Which other subjects (apart from science) do you teach?

a) English

b) Math

c) Hindi

d) Social Science

e) Other

Q8. How many years of work experience as a science teacher do you have, regardless of whether you worked full-time or part-time?

a) Less than 1 year

b) 1–2 years

c) 3–5 years

d) 5–7 years

e) More than 7 years

Q9. Was teaching your first choice as a career? *(A "career" is having a paid job that you regarded as likely to form your life's work.)*

a) Yes

b) No

Module C : Motivation for Entering Teaching

This module elicited teachers' motivations for choosing teaching as a profession, including job security, career stability, work–life balance considerations, intrinsic motivation related to working with children, and family influence.

Q10. Were any of the following factors influenced your decision to become a teacher? *[Multiple select allowed]*

i. Teaching offered a steady career path.

ii. Teaching was a secure job (reliable income).

iii. The teaching schedule (e.g., hours, holidays, part-time positions) fit with responsibilities in my personal life.

iv. Teaching allows me to influence the development of children and young people.

v. It is a respectable profession.

vi. I was inspired by my parents (or other close family member) to become teacher.

vii. None of the above.

viii. Other.

Module D : Teaching Practices and Self-Assessment

Teachers reported the extent to which they engage in specific instructional practices, including motivating students, encouraging critical thinking, using experiments, providing alternative explanations, and supporting learning with digital tools. Items (i)–(vi) use a 4-point scale (*Not at all* to *A lot*); items (vii)–(ix) use a 4-point agreement scale (*Agree a lot* to *Disagree a lot*).

Q11. In your teaching of Science, to what extent can you do the following?

Scale for items (i)–(vi): (a) Not at all (b) To some extent (c) Quite a bit (d) A lot

(i) Get students to believe they can do well in school work.
(ii) Motivate students who show low interest in school work.
(iii) Help students think critically.
(iv) Use science-related experiments as an alternative teaching methods.
(v) Provide an alternative explanation, for example when students are confused.
(vi) Support student learning through the use of digital technology (e.g., computers, tablets, projector).

Scale for items (vii)–(ix): (a) Agree a lot (b) Agree a little (c) Disagree a little (d) Disagree a lot

(vii) I am good at explaining science.
(viii) I do a variety of things (e.g., conducting small experiments, providing examples, etc.) to help my

students learn.
(ix) I explain a topic again when my students don't understand.

Module E : Assessment Practices and Parent Interaction

This module collected information on the frequency of student assessments and the nature of interactions with parents, including participation in parent–teacher meetings and targeted outreach to parents of low-performing students.

Q12. How often do you conduct small tests to grade/mark/rank students?

a) Monthly 1 small test

b) 1 test in 2 months

c) 1 quarterly tests

d) 2 tests in a year (one half-yearly and one final)

Q13. How often do you meet with parents of students?

a) In a usual parents-teacher meeting

b) I sometimes call parents of student if he/she is not doing well academically (or otherwise)

c) We never have any parents-teachers meeting

Module F : Perceptions of Student Performance and Interest

Teachers rated students' performance and interest in science separately for Grades 6, 7, and 8 using ordered categorical scales. These measures capture teachers' perceptions of student engagement and learning and are analyzed descriptively or standardized where appropriate.

Performance ratings (Q14–Q16): applicable only if the teacher teaches the respective grade.

Q14. Grade 6 **Q15.** Grade 7 **Q16.** Grade 8 — How would you rate students' performance/knowledge in science?

a) Excellent

b) Very good

c) Good

d) Not so good

e) Students need to work a lot to improve the standard

Interest ratings (Q17–Q19): applicable only if the teacher teaches the respective grade.

Q17. Grade 6 **Q18.** Grade 7 **Q19.** Grade 8 — How would you rate students' interest toward science on average?

a) Very much

b) Somewhat

c) Very little

d) Not at all

Module G : Job Satisfaction

Teachers reported their overall level of job satisfaction using an ordered categorical scale. This measure is used descriptively and to assess whether the intervention affected reported satisfaction with teaching.

Q20. Are you satisfied with your job?

a) Not at all

b) To some extent

c) Quite a bit

d) A lot

Response Scales and Index Construction in Teacher Survey

Unless otherwise noted, teacher survey items were recorded using ordered categorical response scales, with higher values corresponding to higher levels of the underlying construct. For empirical analysis, selected items related to teaching practices and self-assessment (Module D) are standardized and aggregated into composite indices, while other measures are used individually for descriptive analysis and robustness checks.

E.4 Social Desirability Score

To assess respondents' tendency to provide socially desirable rather than truthful responses, we use a short-form version of the Marlowe–Crowne Social Desirability Scale originally developed by Crowne and Marlowe (1960) and subsequently adapted into reliable and valid short forms by Hays et al. (1989). The scale consists of 5 items and was administered as part of the student survey at the endline. Each item is presented as a statement with a 5-point true/false response scale. For each item, the extreme response corresponding to the socially desirable answer is coded as 1 and all other responses as 0. The social desirability score is constructed by summing the number of socially desirable responses across all items, yielding a total score ranging from 0 to 5. Higher scores indicate a greater tendency toward socially desirable responding, while lower scores indicate a lower propensity for social desirability bias. This measure is used in robustness analyses to assess whether estimated treatment effects on self-reported outcomes are sensitive to differential response tendencies across treatment and control groups.

Response scale (all items): (a) Definitely true (b) Mostly true (c) Don't know (d) Mostly false (e) Definitely false

i. I am always courteous even to people who are disagreeable.

ii. There have been occasions when I took advantage of someone.

iii. I sometimes try to get even rather than forgive and forget.

iv. I sometimes feel resentful when I don't get my way.

v. No matter who I'm talking to, I am always a good listener.