Can Innovation Be Taught In Schools? Experimental Evidence From India

Saloni Gupta[†]

Working Paper

March 20, 2025

Updated Frequently. Click here for the latest version

Abstract

Can innovative ability be taught, or is it primarily inherent? I address this question through a large-scale clustered randomized controlled trial where 5,000 8th-graders in India received either a two-year innovation training program or only opportunities to demonstrate their inherent abilities. Through standardized assessments, expert evaluations, and market validation, I find the program improves innovative ability, problem-solving, and fluid intelligence (0.09-0.19 SD). In a shark-tank style competition, treated teams secured 24 p.p. more funding than control teams. Treatment effects vary by baseline performance: above-median students gain in both higher-order skills and math (0.17 SD), while below-median students improve in higher-order skills but decline in math interest and performance (-0.11 SD). These differential effects suggest students struggling with traditional mathematics redirect effort toward applied contexts where they experience greater success, showing significant capacity for higher-order thinking when engaged with applied content. My results provide the first evidence that disadvantaged youth can acquire higher-order skills increasingly valued in the workforce that can enable them to potentially drive future productivity gains and technological progress.

JEL Classification: I21, I25, O31, J24, C93

Keywords: Educational Innovation, Human Capital Development, Experimental Economics, Skill Formation, Educational Interventions

I am extremely grateful to the Telangana Social Welfare Residential Educational Institutions Society, Telangana Tribal Welfare Residential Educational Society, Inqui-Lab Foundation, and DAI Advisory, for their leadership and collaboration in this study. I especially thank Alejandro Ganimian, Alex Eble, Judith Scott-Clayton, Anjali Adukia, Andreas de Barros, Carolina Concha-Arriagada, Ashutosh Bhuradia, Sandra Black, Abhijeet Singh, Michael Best, and Sarah Cohodes for their feedback and support at multiple stages of this project. I have also benefited from presenting this work at numerous academic venues, including Stanford University, Brown University, Ashoka University, University of Zagreb, Helsinki Graduate School of Economics, Paris School of Economics, University of Chicago, Columbia University, and various Association for Education Finance and Policy conferences. This project is financially supported by NAEd/Spencer Dissertation Fellowship, the Weiss Fund at UChicago, and multiple small research grants at Teachers College. The RCT is registered with the AEA Trial Registry (RCT ID: AEARCTR-0009696).

[†]salonig@stanford.edu, Stanford King Center on Global Development, Stanford University

1. Introduction

Developing countries represent 85% of the global population but contribute less than 15% to global innovation output (WIPO, 2022; UNESCO, 2018). This innovation gap has substantial economic implications, as Jones (2022) demonstrates that developing countries could quadruple innovation output through improved talent allocation and education. While Bell et al. (2019a) show that early innovation exposure significantly determines who becomes an inventor, causal evidence on whether innovation skills can be systematically taught remains scarce—particularly in developing countries where potential returns are largest.

Over the past decade, numerous developing countries have implemented early innovation education programs in schools, providing an ideal setting to address this gap.[1] These programs break from traditional educational STEM models by integrating hands-on experimentation and science-based tools to develop the skills to create or significantly improve existing products that could add value to end users—aligning with the OECD (2018) definition of innovation. Such programs could affect economic outcomes through two primary channels. First, they can develop future innovators who can drive technological advancement. Second, they could build skills relevant for the approximately 88% of workers in developing economies employed in the small enterprise sector (ILO, 2018). For these workers, innovative abilities can help identify inefficiencies and improve technological adoption. Evidence shows that addressing such readily achievable inefficiencies yields productivity gains of 8-30% (Bloom et al., 2013; Jensen and Miller, 2018; Adhvaryu et al., 2018; Atkin et al., 2017).

In this paper, I present experimental evidence on the impact of a technology-led innovation education program on students' higher-order skills, academic outcomes, and career aspirations. The study was implemented in 80 residential schools in Telangana, India, which were randomly assigned to either treatment (40 schools) or control (40 schools) conditions. These schools admit exclusively students from constitutionally protected marginalized communities (Scheduled Castes and Scheduled Tribes), providing evidence on whether innovation skills can be developed among populations historically underrepresented in innovation careers. I follow approximately 5,000 students from Grade 8 through Grade 9 during academic years 2022-2024.

A key feature of this study is the development of new measurement tools for assessing economically valuable higher-order skills in resource-constrained environments. Drawing

[1] This trend is documented across multiple reports and studies (World Bank, 2016; UNESCO, 2017; OECD, 2015), with notable implementations in India (NITI Aayog, 2020), Rwanda (Ministry of ICT and Innovation, Rwanda, 2019), Thailand (Ministry of Education, Thailand, 2018), and several Latin American countries (Blikstein et al., 2022).

on approaches from both economics and cognitive psychology, I develop and validate measures that function effectively without requiring sophisticated technology infrastructure. The innovation measures include a novel innovation quality scale calibrated against practicing innovators, an adaptation of the multi-armed bandit task from Ederer and Manso (2013a), and offline computer-based instruments for assessing fluid intelligence, grit, emotional intelligence, and problem-solving skills.

The intervention was implemented by Inqui-Lab Foundation, an NGO founded by engineers-turned-teachers who developed a specialized pedagogy for underserved populations that mirrors curriculum used in leading engineering universities. The program delivered 60 hours of training through 40 weekly sessions of 90 minutes each week in 40 treatment schools over two years. Each school received standardized kit boxes containing circuits, sensors, and recycled materials valued at Rs. 15,000 ($180 USD) per classroom. The program implemented four sequential 8-session blocks focused on developing specific innovation skills: understanding existing technologies, identifying local problems, creating and testing prototypes, and refining solutions based on feedback. Students progressed through a defined sequence where early sessions utilized directed critique discussions of existing solutions (e.g., analyzing the design evolution of everyday objects), middle sessions employed guided problem-finding activities (e.g., school walkabouts to identify sustainability or local community challenges), and culminating sessions focused on team prototype construction using circuits and sensors to address identified problems. Two student leaders facilitated each classroom, participating in debrief circles after each curriculum block to standardize implementation.

Control schools received only information about an annual innovation challenge with detailed criteria but no structured training, similar to statewide challenges previously conducted by the Telangana government (Government of Telangana, 2019; NITI Aayog, 2021a). This created a research design that tests whether structured training improves innovation outcomes compared to inherent abilities alone.

I report three key findings. First, I find that the program significantly improved students' innovation skills consistently across three different measures. My first measure uses a high-risk, high-reward multi-armed bandit problem adapted from Ederer and Manso (2013a) and grounded in March's (1991) innovation framework. I find that treated students exhibited a greater inclination towards exploratory behavior (0.127 SD, $p < 0.01$) and discovered higher value-added solutions (0.18 SD, $p<0.05$). For my second measure at the team level: treated groups produced higher quality ideas (0.243 SD, $p < 0.01$) with greater perceived funding potential (0.177 SD, $p < 0.10$), as measured by a novel idea quality metric calibrated against practicing inventors. My third measure captures real-world impact: treatment teams secured 24% more funding than control teams in an

inter-school shark tank competition. All these innovation-related effects remain statistically significant after adjusting for multiple-hypothesis testing.

Second, I find improvements in fluid intelligence (0.129 SD, $p < 0.10$) and problem-solving (0.191 SD, $p < 0.05$) with minimal impact on emotional intelligence (0.081 SD, $p > 0.10$) and grit (-0.067 SD, $p > 0.10$).

Third, I find opposite effects on math by baseline ability: gains for high-performers (+0.18 SD, $p < 0.05$) but declines for struggling students (-0.12 SD, $p < 0.05$). I find no significant average effects on math (0.066 SD, $p > 0.10$) and science test scores (0.065 SD, $p > 0.10$) across the full sample. But for students with above-median baseline math scores, the innovation program had complementary effects, improving both innovation skills and math performance. These students also showed increased STEM career aspirations. In contrast, for students with below-median baseline performance, the program improved innovation and problem-solving skills (0.12-0.18 SD) but reduced math motivation (-0.09 SD, $p < 0.10$) and scores (-0.12 SD, $p < 0.05$) While these individual effects are statistically significant, they do not maintain significance when adjusting for multiple hypothesis testing.

This study provides the first experimental evidence on teaching innovation skills to disadvantaged youth, contributing to two primary strands of literature in economics.

First, I contribute to the literature on higher-order skill development by providing rigorous evidence on a bundled innovation training program in a developing country context. Human capital research shows higher-order skills yield increasing returns as technology transforms labor markets (Deming, 2022; Acemoglu and Autor, 2011; Cunha and Heckman, 2007). Recent studies demonstrate that targeted programs can develop metacognitive abilities (Alan et al., 2019a; Alan and Ertac, 2018; Yeager et al., 2019; Alan and Mumcu, 2024; Bettinger et al., 2018; Alan and Ertac, 2018), negotiation and business skills (Chioda et al., 2021; Ashraf et al., 2020b), and problem-solving skills (Bando et al., 2019; Ashraf et al., 2020a), with effects typically ranging from 0.1-0.4 standard deviations. I extend this literature by examining an intervention that simultaneously targets multiple skills critical for innovation in a developing country context.

I also develop new measurement tools that address a key challenge in studying higher-order skills in resource-constrained environments. While economic theory recognizes non-cognitive skills as crucial determinants of productivity (Deming, 2022; Cunha and Heckman, 2008), measuring these skills remains difficult, especially in developing countries. My study develops and validates instruments for measuring innovation skills, fluid intelligence, emotional intelligence, and problem-solving that function effectively in resource-constrained environments, enabling more widespread inclusion of these economically valuable skills in educational evaluation.

Second, I extend the innovation literature by examining who develops innovation skills and what problems they address. Research shows childhood exposure (Bell et al., 2019a), access to knowledge (Biasi and Ma, 2022; Toivanen and Väänänen, 2016), and family background (Aghion et al., 2017a) shape who becomes an inventor, creating substantial underrepresentation of disadvantaged groups. This matters because inventor demographics influence which problems receive attention—women inventors focus more on women's health (Koning et al., 2021) while inventors generally address problems familiar from their socioeconomic experiences (Einiö et al., 2023). The resulting innovation gaps disproportionately affect developing countries, where lost innovative potential may reduce output by up to 40% (Jones, 2022). My study demonstrates that marginalized students not only develop innovation skills but create solutions specifically addressing challenges in their local contexts, providing evidence that diversifying innovation participants changes which problems receive attention—even at the educational stage.

Beyond economics literature, my findings expose how India's education system pursues two contradictory goals: promoting higher-order skills while sorting students through memorization-based examinations—a fundamental tension that undermines educational equity. While the National Education Policy 2020 (India's Ministry of Education, 2020) and Atal Tinkering Labs (NITI Aayog, 2021a) advocate for innovation and higher-order thinking, they operate within a system dominated by memorization-focused assessments. In classrooms, 90% of teacher questions require simple recall (Bhattacharjea et al., 2011), with instruction centered on standardized test preparation. Muralidharan (2024) argues this approach stems from education's primary function as a sorting mechanism for limited higher education opportunities rather than skill development. This emphasis on exam culture driving classroom instruction creates a divide between academic and applied skills, consistent with Banerjee et al. (2025)'s finding that students can expertly solve calculation problems in real-world contexts yet struggle with identical math concepts in academic formats requiring memorized procedures.

I show that while all students developed higher-order skills, the math gap between them widened. High-performing students effectively transferred skills across domains, while lower-performers engaged more with hands-on tasks and reallocated resources away from abstract math—evidenced by their reduced math motivation. This heterogeneous response is consistent with students making rational choices about effort allocation based on their comparative advantages (Becker, 1965). When presented with alternative learning opportunities, students who struggle with traditional mathematics may rationally redirect effort toward innovation activities where they experience greater immediate success and engagement. These findings demonstrate that academically struggling students can learn valuable higher-order skills, despite being disadvantaged by a system that ulti-

mately prioritizes memorization over the very capabilities its policies claim to value.

The remainder of the paper is organized as follows. Section 2 summarizes the key features of the program and the context in which it was implemented. Section 4 details the evaluation design. Section 5 gives a detailed account of the toolkit I use to evaluate the program. Section 6 describes the data. Section 7 presents my main results. In Section 8, I discuss the mechanisms through which the program might have improved higher-order skills and caused divergence in math achievement for the treatment group. In Section 10, I discuss cost-effectiveness and potential scale-up issues, and I conclude in Section 11.

2. Context, Intervention, and Delivery

2.1 Context

India faces a fundamental human capital allocation challenge: despite rapid growth in high-skill sectors, most workers remain in low-productivity traditional industries. Only 7% of young people complete higher education (AISHE, 2020), while 88% of workers enter small enterprises (ILO, 2018).

This situation stems from three interconnected crises. India confronts a severe learning crisis in basic education, with only 50% of fifth-grade students able to read at a second-grade level (ASER, 2022). The country simultaneously faces persistent youth unemployment, reaching 23.2% among those aged 15-29 (PLFS, 2022). Even those who complete technical education face the third crisis: an employability gap where 80% of engineering graduates require significant additional training before being "employable" (India Skills Report, 2022), reflecting a disconnect between learning and workforce needs.

India's economic goals are ambitious but human capital development is not keeping pace with these economic challenges. Initiatives like Make in India, Startup India, and the National Artificial Intelligence Strategy assume the availability of innovation-capable workers across sectors. However, the traditional education system—focused primarily on rote learning and standardized testing—struggles to develop higher-order thinking skills these economic initiatives require.

To address this gap, several innovation-focused educational initiatives have emerged. The Atal Innovation Mission has established over 10,000 Atal Tinkering Labs in schools nationwide with another 10,000 in process, representing a $240 million investment aimed at creating "one million young innovators" (NITI Aayog, 2021a). Delhi's Entrepreneurship Mindset Curriculum ($27 million annually) targets similar skills at scale. Despite these investments, little causal evidence exists on their effectiveness.

Telangana exemplifies India's economic duality: Hyderabad's tech sector generates

30% of state GDP with just 7% of workers, while small enterprises (38% of workers, 23% of GDP) and agriculture (55% of workers, 18% of GDP) operate at far lower productivity levels (Government of Telangana, 2022). The state has responded with its 2015-2025 Innovation Policy, an annual School Innovation Challenge engaging 3,000+ schools, and infrastructure like T-Works, the state-sponsored prototyping center and makerspace supporting hardware innovation and skills development (Government of Telangana, 2019).

These innovation education efforts could particularly benefit marginalized communities by developing applied thinking skills relevant across economic sectors. The residential schools in this study serve exclusively Scheduled Castes and Scheduled Tribes—constitutionally protected groups comprising 24.5% of Telangana's population but representing less than 5% of technology sector employment. These communities face persistent barriers to economic mobility (Asher et al., 2022; Hnatkovska et al., 2013) and may benefit from skills that enhance productivity in traditional sectors while potentially opening pathways to higher-value economic activities.

This research examines a structured innovation education program at a critical juncture—8th and 9th grades—when Indian students begin specializing before high-stakes examinations that determine their economic trajectories. The study provides initial evidence on whether such programs can develop economically valuable skills among students historically excluded from India's high-productivity sectors.

2.2 Intervention and Delivery

The Inqui-Lab innovation education program followed the same cohort of approximately 2,500 students from grade 8 through grade 9 across 40 government residential schools in Telangana state in years 2021-2023.

The curriculum followed a design thinking framework comprising five defined phases: Find, Learn, Ideate, Make & Test, and Present. This approach aligns with Razzouk and Shute (2012) conceptualization of design thinking as "an analytic and creative process that engages a person in opportunities to experiment, create and prototype models, gather feedback, and redesign." Throughout the two-year program, students completed this innovation cycle multiple times with increasingly difficult challenges, participating in over 40 weekly sessions of 90 minutes each totaling approximately 60 hours of instructional time.

A key feature of the intervention was its self-directed learning approach. Students worked primarily from four workbooks of about 170 pages each that provided step-by-step instructions for each activity, enabling teams to progress through the curriculum with no adult supervision. Student leaders managed timing and facilitated discussions but were instructed to answer only clarifying questions rather than providing solutions. This

designed independence ensured that teams developed innovation skills through their own discoveries and collaborations, with the workbooks sequentially organizing the learning progression across sessions.

The program developed innovation skills through four sequentially organized phases with distinct learning objectives:

In the initial phase, students explored the evolution of everyday objects, examining historical iterations of items like pens, earphones, and televisions to understand product development principles. They learned to identify design improvements through documented observation techniques, documenting changes in functionality, materials, and user experience. Students then applied this understanding through the full innovation cycle, ultimately redesigning everyday school objects like chalk holders and specialized dusters to address needs they identified among their peers and teachers.

As students progressed through the basic design thinking methodology, they incorporated more complex content by exploring sensor technologies and basic electronics. Still following the same five-step process from need identification to presentation, teams built working prototypes incorporating light sensors, temperature detectors, and motion sensing components. This iteration resulted in operational innovations such as a health monitoring system for cardiac patients, a safety device linking helmets with motorcycle operation, an automatic farming machine using solar power, and an efficient water distribution system for irrigation.

After being introduced to both the design methodology and technical skills, students focused more deeply on environmental sustainability issues. They conducted systematic observations and user interviews in their school environments, applying multiple data collection research techniques. Teams documented measurable needs through methods like waste audits in school kitchens, water usage mapping, and electricity consumption tracking, developing data-informed solutions to address resource conservation challenges.

The program ended with independent capstone projects where students integrated all previous skills to address challenges of their choosing. These capstone projects were documented by the research team at the endline.

The intervention employed a peer-to-peer facilitation model where two student leaders per classroom coordinated activities and facilitated peer feedback. These leaders, nominated by teachers, received 20 hours of additional training at the beginning of the year and ongoing teacher support. Students worked in collaborative teams of 4-5 members, using innovation kits containing developmentally appropriate materials: simple circuit components, stationery supplies, recyclable items, and beginner-level tools suitable for middle school students.

The workbooks integrated timed reflection throughout the innovation process. At the

end of each session, students completed individual reflection prompts about their contributions and learnings, followed by team discussions about their collaborative effectiveness. These reflections included prompts like "What one thing will our team do differently next time?" and "How well did we listen to each other's ideas today?" Teams also rated their progress on idea development, prototype refinement, and presentation skills, setting measurable improvement goals for subsequent sessions. This reflective practice was reinforced through periodic peer feedback activities where teams exchanged actionable suggestions and identified strengths in each other's work.

To ensure treatment fidelity, seven trained observers conducted over 400 classroom observations throughout the two-year period using standardized protocols. Implementation quality was assessed through standardized observation forms tracking completion of prescribed activities, facilitation techniques, material usage, and student engagement. The monitoring data showed 85% of observed sessions fully implemented all required program elements. Additionally, a randomly selected subset of student teams was followed longitudinally to document their progression through the complete innovation process, providing detailed qualitative data on skill development over time.

This monitoring approach verifies uniform implementation across all treatment schools, strengthening confidence in the internal validity of the intervention.

3. Evaluation Design and Timeline of the Study

I selected 80 schools from a pool of 428 eligible schools based on operational feasibility for the implementation partner and input from the state government ensuring no other related programs were occurring in these schools during the study period. Within this selected group, I employed a stratified randomization design. Stratification was based on three key dimensions: geographical setting (urban/rural), gender composition (girls'/boys' schools), and administrative affiliation (social welfare/tribal welfare departments). I randomly assigned 40 schools to the treatment group and 40 to the comparison group within each stratum.[2]

The intervention consisted of structured innovation training provided to the treatment group on a weekly basis. The comparison group did not receive this training. Instead, at the beginning of each academic year, comparison schools were formally notified about an upcoming 'Shark Tank Challenge' via an official letter. This communication emphasized

[2]An a priori power analysis was conducted for this cluster-randomized trial. With 80 schools and an assumed average of 50 students per school, I calculated 80% power to detect an effect size of approximately 0.15 standard deviations at the 5% significance level (two-tailed), assuming an intraclass correlation coefficient of 0.15 and that baseline covariates would explain about 30% of the outcome variance. Full details of the power calculation are provided in Appendix **??**.

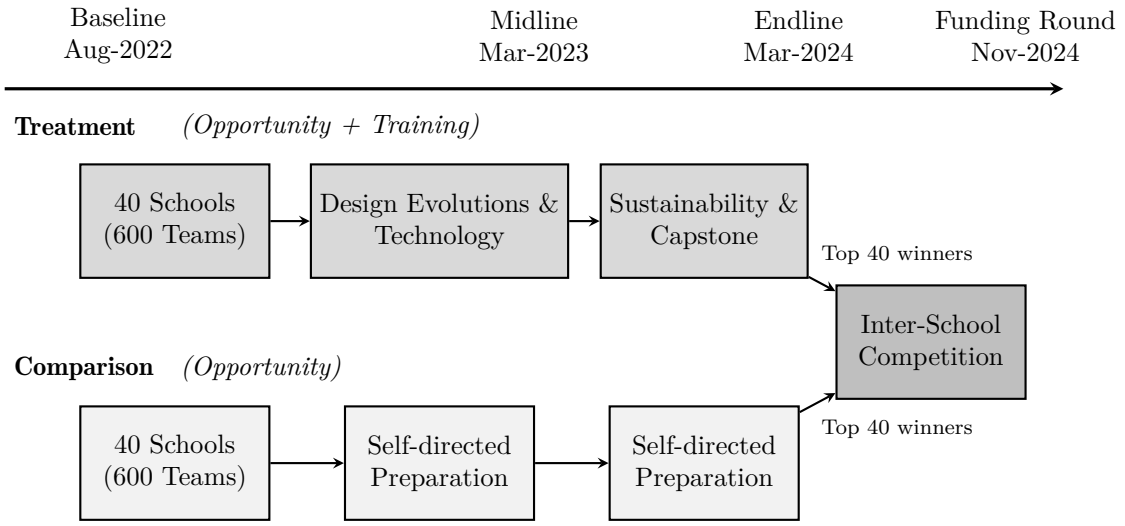


FIGURE 1: PROGRAM TIMELINE

Notes: The timeline illustrates that 80 schools were divided into treatment and comparison groups using strata randomization. In every school, students were split into fifteen teams, with each team consisting of 4 students chosen at random. This procedure is repeated in comparison schools as well. Treatment schools received structured design thinking curriculum delivered through workbooks and peer facilitators. Comparison schools had the same opportunity to develop ideas for the competition but without specialized training. Both groups consist of similar residential government schools.

that teams would be evaluated based on how beneficial their innovative idea was for its intended users. The letter also provided basic competition logistics but offered no guidance on idea development or preparation. This innovation challenge concept was familiar to students due to previous state-organized innovation competitions (Telangana State Innovation Cell, 2021; AMS Public Relations Society, 2022) and the widespread popularity of entrepreneurship reality shows in India (Kushagra, 2023; Soni and Sharma, 2022; Wadhwa, 2022).

As shown in Figure 1, both treatment and comparison schools participated in the same competition timeline from August 2022 to November 2024, with baseline, midline, and endline assessments. The key difference was in their preparation approach. While treatment schools received 40 structured sessions totaling 60 hours of formal training over two years, comparison schools engaged in "Self-directed Preparation" alongside the standard school curriculum. Since these were residential schools, students in both groups had independent homework periods after regular school hours. Treatment schools scheduled their weekly 90-minute sessions during independent study periods, ensuring no academic subject time was displaced. Comparison group teams were encouraged to use this same time to develop their innovation ideas on their own. Thus, both groups utilized the same independent study time allocation, but the treatment group received structured training while the comparison group engaged in self-directed preparation. Ultimately, top performers from both groups would compete in the annual Inter-School Competitions, which were organized in October 2023 and November 2024.

Both groups were thus motivated by the prospect of a competition, but only the treatment group received structured training. This design allows me to assess whether formal innovation training leads to better outcomes compared to students' self-directed efforts in preparing for a familiar type of innovation challenge.

The experiment spans two academic years, with data collection at three primary points: baseline in August 2022, midline in March 2023 (end of Year 1), and endline in March 2024 (end of Year 2). An inter-school competition for winning teams from each school was conducted in November 2024, in the following academic year, due to logistical constraints and school calendar considerations. The map in Figure **??** shows the geographic distribution of treatment and comparison schools across 12 districts in Telangana state.

4. Outcomes

My primary objective is to estimate the causal impact of the program on students' innovative abilities, as well as related higher-order skills[3] including problem-solving skills, grit, emotional intelligence, and fluid intelligence. The program did not directly target achievement outcomes. However, I measured performance in mathematics and science to examine whether the development of higher-order skills generated positive spillovers to academic domains.

4.1 Innovative Ability

Recognizing the challenge of measuring innovative abilities in young students, I employ three distinct measures at individual and team levels. First, I use a lab-in-the-field game based on March (1991b)'s seminal work on exploration and exploitation in innovation processes, using a multi-arm bandit problem to assess individual innovative behavior. Second, all students collaborate in teams to submit ideas, which are evaluated using an idea quality measure developed with a survey of experienced innovators. This measure has established predictive validity for real-world innovation outcomes such as user reach and funding raised. Third, the top-performing teams, selected based on their idea quality, participate in an inter-school funding competition. This competition features expert jury assessment, including non-profit entrepreneurs, social impact investors, and corporate innovators, employing a double-blind process with two jurors evaluating every 20 teams, balanced across treatment and control groups.

[3]There are various terms used in the literature to describe non-academic abilities, including 'non-cognitive skills,' 'soft skills,' and 'socio-emotional skills.' Following Deming (2022), I use the term 'higher-order skills' to emphasize that these abilities involve complex cognitive processes that extend beyond basic academic knowledge. This terminology also avoids the misleading cognitive/non-cognitive dichotomy since higher-order skills often involve both cognitive and socio-emotional components

4.1.1 Lab-in-the-field-game using multi-armed bandit problem

This study employs a multi-armed bandit problem to measure innovative ability in students. I ground this approach in March (1991b)'s exploration-exploitation framework, building upon a growing body of literature that uses multi-armed bandit tasks to measure innovative behavior (Ederer and Manso, 2013b; Herz et al., 2014; Bandiera et al., 2021).

March's framework defines innovative ability through the interplay of exploration and exploitation behaviors. Exploration involves the willingness to search for and experiment with new possibilities that might yield superior outcomes, despite uncertainty and higher risk of failures. Exploitation is the ability to recognize and refine successful discoveries to extract their maximum value. This behavioral signature suggests that innovation can be measured by observing how individuals navigate scenarios requiring these trade-offs.

Neuroscientific studies provide insights into the neural basis of these behaviors. Daw et al. (2006) and Laureiro-Martínez et al. (2015) show that exploration in multi-armed bandit problems engages cognitive systems associated with control and decision-making under uncertainty, while exploitation activates systems linked to immediate reward processing. These distinct neural systems support March's assertion that balancing these processes is crucial for adaptive learning and innovation.

The multi-armed bandit task is administered through the offline Qualtrics app on individual tablets. I chose a business innovation challenge with common Indian street businesses, contexts familiar to students through everyday observations. This design deliberately differs from the innovation program's curriculum to minimize potential biases while ensuring all participants have a similar frame of reference.

Students engage with this problem over 12 distinct periods, choosing between business ventures with different risk-reward profiles. The Fresh Juice Stand (exploitation arm) offers more certain but lower rewards with a 25.7% chance of zero profit and maximum profit of 110. The Ice Cream Stand presents a more extreme trade-off with a 63.3% chance of zero profit but a maximum profit of 250. After each period, students receive feedback, allowing them to learn from their decisions, mirroring real-world innovation scenarios.

The task measures three key aspects of innovative ability: (1) propensity to be exploratory initially, (2) number of times students choose to explore as well as exploit in all periods, and (3) success in identifying high-value outcomes. These metrics collectively capture an individual's ability to discover solutions superior to the default option.

4.1.2 Idea Quality

Building on the multi-armed bandit game, I developed a complementary measure to address the limitations in capturing real-world innovation ability. I created the Innovation-S scale, a novel psychometric tool designed to measure the quality of innovative ideas.

The Innovation-S scale was developed through a process involving in-depth interviews with 14 successful innovators, including founders, scientists, and a Nobel Laureate, followed by a survey of 44 experienced innovators from diverse fields. The participants represented a range of industries, with a significant concentration in education (40.9%) and technology (15.9%), and included founders (25%), product managers (13.6%), and analysts (11.4%) among others. This diverse group had experience across multiple innovation stages - 66% in ideation, 61% in development, 55% in launching, and 59% in improvement, providing a balanced perspective across different organizational structures.

The resulting scale has a 2-factor structure (novelty and value-add) comprising seven items, such as "offers a new solution to a problem" and "improves the lives of those who will use it." The scale demonstrates valid psychometric properties (Cronbach's alpha = 0.69) and strong predictive validity, with correlations of 0.71 and 0.31 with user base and funding respectively. These metrics indicate that the Innovation-S scale is a reliable tool for measuring innovative ability, capable of predicting real-world innovation outcomes across various scales of operation and funding levels.

To apply this scale in evaluating student ideas, I carefully selected and trained two expert evaluators from a pool of over 1,200 applicants. These evaluators independently rated each idea submitted by both treatment and comparison teams using the Innovation-S scale and provided a "willingness to fund" measure for each idea.

4.1.3 Interschool Fundraising Challenge

To assess the practical impact of the innovation training program, I implement a real-world fundraising challenge as the third outcome measure. From the pool of ideas collected at the end of the year, I selected the top 40 ideas from each group based on the raters' ratings, resulting in 80 total ideas for the challenge. To minimize preparation bias and simulate real-world constraints, participants receive only 5 days' notice before the challenge.

The evaluation process involves an expert jury comprising eight individuals: non-profit entrepreneurs, social impact investors, and corporate innovators. I employ a double-blind assessment method, where each pair of jurors evaluates 20 teams, balanced across treatment and control groups. The key outcome measure is the difference in average fundraising amounts between the treatment and control groups.

4.2 Higher Order Skills

I measure four higher-order skills using adapted versions of validated instruments: fluid intelligence using Raven's Advanced Progressive Matrices test (Raven, 1938), emotional intelligence using the "Reading the Mind in the Eyes" test (Baron-Cohen et al., 2001), grit using tool developed by (Alan et al., 2019b), and problem-solving using a set of five Sudoku puzzles (Yato and Seta, 2003).

4.2.1 Problem-Solving Ability

Problem-solving and innovation represent distinct yet complementary cognitive processes. Problem-solving involves working within established parameters to find optimal solutions to well-defined problems with known constraints, clear goals, and available methods for solution (Newell and Simon, 1972; Russell and Norvig, 2010), while innovation requires navigating what Simon (1973) characterizes as ill-defined problems—situations where the goals are ambiguous, constraints are unclear, and solution paths must be generated rather than merely discovered. This distinction aligns with March (1991b)'s framework, where innovation requires exploration of uncertain alternatives with incomplete information about potential solutions.

For problem-solving assessment, I use 4×4 Sudoku puzzles, which maintain the essential structure of NP-complete problems (Yato and Seta, 2003)—computational challenges where solutions can be verified quickly (in polynomial time) but may require systematic exploration to discover. Students completed five timed puzzles (2 minutes each) on tablets using an offline application calibrated for grade 8 students. This format was specifically calibrated for grade 8 students, providing an age-appropriate measure of logical reasoning, pattern recognition, and constraint satisfaction strategies—key components of problem-solving ability within a well-defined rule set.

By measuring both abilities, I can evaluate whether the intervention differentially affects problem-solving (the ability to efficiently find solutions within well-defined constraints) versus innovation (the capacity to generate novel value under uncertainty)—skills that may develop through different cognitive pathways.

4.2.2 Fluid Intelligence

Fluid intelligence plays a crucial role in innovation (Carpenter et al., 1990; Jauk et al., 2013). Recent economic literature underscores the importance of these cognitive abilities in driving innovation and economic growth, with Aghion et al. (2017b) and Toivanen and Väänänen (2016) both demonstrating strong positive correlations between IQ measures and innovation outcomes, showing that higher cognitive ability predicts both the quantity

and quality of patents.

To measure fluid intelligence, I employ Raven's Progressive Matrices, a widely recognized and culturally fair assessment that minimizes the influence of language skills and prior knowledge, reducing cultural and educational biases (Raven, 2000).

I expect that the intervention will improve fluid intelligence in adolescents by leveraging their developmental stage and engaging them in innovative thinking. For instance, during the intervention, students study recent technological advancements and conceptualize them to adapt to locally-relevant innovations. This process of recognizing patterns, adapting technologies, and solving real-world problems likely accelerates fluid intelligence development beyond typical adolescent growth.

4.2.3 Emotional Intelligence

Emotional intelligence plays a critical role in the landscape of innovation, increasingly characterized by collaboration and teamwork. Recent studies document a significant trend towards collective knowledge production across science (Jones, 2009; Wuchty et al., 2007; Jaravel et al., 2018) and highlight the importance of social interactions in knowledge diffusion (Akcigit et al., 2018).

A typical school setting in India is largely characterized by lecture-style teaching with minimal teacher-student interaction and limited opportunities for active learning (Djaker et al., 2024). In stark contrast, this intervention incorporates several components that could potentially leverage adolescent neuroplasticity to improve emotional intelligence, including intensive teamwork, peer leadership, structured emotional reflection, and regular presentation opportunities.

To measure emotional intelligence, I used the Reading the Mind in the Eyes Test (RMET), developed by Baron-Cohen et al. (2001). This test assesses an individual's ability to recognize and interpret emotions from photographs showing only the eye region of faces, serving as a quantitative measure of emotional recognition and empathy (Vellante et al., 2013). The adaptation used in this study maintains the core assessment principles while ensuring cultural relevance, as detailed in Appendix **??**.

My conjecture is that these intervention features may foster specific dimensions of social emotional development—particularly empathic accuracy, perspective-taking, and emotional recognition—in ways that traditional educational approaches do not typically address (Alan and Kubilay, 2025). However, the effects may not be uniformly positive. Given students' limited exposure to social skill development, some may thrive in this new collaborative environment, while others might struggle with the unfamiliar interpersonal dynamics of team-based innovation work.

4.2.4 Grit

Grit, characterized by perseverance and passion for long-term goals, plays a crucial role in innovation processes. Thomas Edison's relentless pursuit of a functional light bulb, famously requiring thousands of attempts, epitomizes grit in innovation. Studies have demonstrated grit's predictive power for success across various domains (Duckworth et al., 2007) including the ability to achieve breakthrough innovations (Galton, 2019; Mooradian et al., 2016; Jachimowicz et al., 2018).

In typical educational settings, opportunities to develop grit are often limited. This intervention incorporates elements of the design thinking process that may foster grit development, with students engaging in finding problems, generating ideas, and building prototypes with limited resources. Each stage presents unique challenges that require persistence.

My hypothesis is that these iterative processes, coupled with the program's long-term nature and team-based approach, may cultivate grit. However, the development of grit may not be uniform across all participants, as some students might resort to easier problems or less challenging solutions due to various factors, especially those who lack intrinsic motivation for the program or the end-of-year competition, given the largely self-directed nature of the learning experience.

To measure grit, I employed a behavioral measure adapted from Alan et al. (2019b), where students choose between easy and hard versions of five puzzles. In this assessment, students are first randomly assigned to either an easy or hard version of the initial puzzle to experience the difference in challenge and potential rewards. For each subsequent puzzle, they must decide whether to attempt an easy version with a high probability of success but lower reward (a standard ballpoint pen), or a difficult version with a lower probability of success but higher reward (a premium gel pen) if completed. This approach captures grit through revealed preferences for challenge rather than self-reported perseverance, providing a behavioral indicator of students' willingness to persist through difficulty for greater rewards. The measure quantifies grit as the proportion of difficult options selected across multiple decision points,

4.3 Academic Outcomes and Motivation

Literature often cites academic ability as predictor of being an innovator (Bell et al., 2019b). To assess potential spillover effects on these domains, I employed standardized assessments for mathematics and science. For mathematics, I used assessments from the Mindspark study (Muralidharan et al., 2019b) and school textbooks, chosen for their reliability and relevance to the Indian educational context. For science, I developed a custom assessment based on a review of school textbooks and workbooks, ensuring

alignment with the curriculum. Both math and science assessments were validated with out-of-sample students to ensure their reliability and appropriateness.

I also collected self-reported measures of attendance, classroom engagement (number of questions asked to teachers), social connections (names of friends in classroom), and career aspirations to examine broader impacts of the program. These measures capture potential mechanisms through which innovation education might influence student outcomes beyond direct cognitive effects. For instance, increased interest in innovation could boost attendance and classroom participation (Yeager et al., 2019), while collaborative activities might expand students' social networks in meaningful ways (Hoxby, 2000; Zárate, 2022; Sacerdote, 2014). Additionally, exposure to innovation processes could reshape career aspirations toward more entrepreneurial or STEM-focused pathways (Bell et al., 2019b). These secondary outcomes provide insight into how the intervention might influence student trajectories through both educational engagement and future orientation channels.

5. Data and Empirical Analysis

Based on official school registry data, the total potential sample included 5,871 registered students across the 80 schools, with approximately equal numbers in treatment and control schools. At baseline (August 2022), I collected data from 4,948 students, representing 84.9% of registered students in the control group and 83.7% of registered students in the treatment group. At midline (March 2023), my sample included 5,077 students, representing 87.9% of registered control students and 85.1% of registered treatment students. By endline (March 2024), I had data from 4,825 students, representing 79.9% of registered control students and 84.5% of registered treatment students.

5.1 Outcome Measures by Data Collection Wave

I collected a wide range of outcome measures across three data collection waves. At baseline (August 2022), I measured mathematics and science achievement, Ravens Progressive Matrices (fluid intelligence), Reading Mind in Eyes Test (emotional intelligence), Big Five personality traits, risk preferences, and innovation task performance (multi-armed bandit task). I also collected data on social networks (number of friends) and career aspirations.

At midline (March 2023), I collected the same set of academic achievement measures (math and science scores), cognitive assessments (Ravens and RMET), innovation task performance, and Big Five personality traits. I also gathered data on subject preferences across eight school subjects, risk preferences, attendance, classroom participation, and updated social network and career aspiration measures.

At endline (March 2024), I expanded the assessment battery to include problem-

solving skills and grit in addition to the previously collected measures of academic achievement, fluid and emotional intelligence, and innovation task performance. I added detailed time allocation measures across various activities (school, homework, sports, innovation, other) and collected qualitative data including creative writing samples, innovation improvement suggestions, and evaluations of external innovation ideas. I also collected one innovation idea from each treatment and control team, resulting in 1,191 ideas for evaluation (564 from comparison and 627 from treatment teams).

To enable covariate adjustment and improve precision, I constructed a socioeconomic status (SES) index using principal component analysis based on eight household asset indicators.

5.2 Internal Validity

Table 1 presents baseline characteristics for schools and students in my sample, comparing treatment and control groups and data collected in each round. As expected from my stratified randomization, treatment and control groups are perfectly balanced on key stratification variables—tribal schools, girls' schools, and rural schools (all $p = 1.00$). Other school characteristics show no statistically significant differences, with the joint F-test ($F = 0.089$, $p = 0.999$) confirming balance at the school level. At the student level, most characteristics show no significant differences, with a small exception in socioeconomic status (difference=0.17, $p = 0.005$). Crucially, my primary outcome measures of innovation skills ($p = 0.800$), math ($p = 0.778$), and science ($p = 0.840$) are well-balanced. The joint F-test ($F = 1.572$, $p = 0.117$) indicates no systematic differences between treatment and control students.

The table also presents sample composition across data collection waves. While attendance patterns show some variation between treatment and control groups across waves ($F = 3.002$, $p = 0.035$), the proportion of students present in all three waves remains similar between groups (58.5% control vs. 59.8% treatment, $p = 0.637$). Additional analyses presented in Table **??** confirm that baseline characteristics remain balanced among students present at endline, with no evidence of differential selection based on baseline achievement ($p = 0.925$).

5.3 Empirical Strategy

To estimate the impact of the innovation training program, I use the following regression specification for individual-level outcomes:

$$Y_{is} = \alpha + \beta \text{ Treatment}_s + \gamma \text{ Strata}_s + \delta Y_{is}^{baseline} + \phi X_{is} + \psi S_s + \varepsilon_{is} \quad (1)$$

where Y_{is} represents the outcome of interest for student i in school s, Treatment_s

Table 1: Balance Table

Variable	All	Treatment	Control	Difference	p-value
Panel A: School Characteristics					
Tribal School	50.00	50.00	50.00	0.00	1.00
Girls School	50.00	50.00	50.00	0.00	1.00
Rural School	62.50	62.50	62.50	0.00	1.00
Library	50.00	50.00	50.00	0.00	1.00
Playground	13.75	15.00	12.50	2.50	0.75
Ramp Facility	70.00	67.50	72.50	-5.00	0.63
Water Facility	1.00	1.00	1.00	0.00	.
Computer Lab	66.25	65.00	67.50	-2.50	0.82
Toilets	61.26	58.12	64.40	-6.28	0.60
Observations	80	40	40		
F-statistic: 0.089		p-value: 0.999			
Panel B: Student Characteristics and Baseline Outcomes					
Age	15.48	15.46	15.49	-0.03	0.643
SES Index	2.33	2.42	2.25	0.17	0.022
Mother's Occupation (Housewife)	0.39	0.40	0.37	0.03	0.426
Father's Occupation (Farmer)	0.68	0.65	0.71	-0.06	0.167
Computer Skills	34.33	35.13	33.54	1.59	0.681
Number of Friends	4.77	4.80	4.74	0.06	0.756
Math Score	41.72	41.24	42.21	-0.97	0.637
Science Score	53.63	53.61	53.64	-0.03	0.984
Fluid Intelligence	58.49	58.31	58.67	-0.36	0.833
Emotional Intelligence	44.32	43.96	44.68	-0.72	0.451
Innovation Skills	45.66	45.97	45.35	0.62	0.683
Risk Preference Index	0.00	-0.02	0.02	-0.04	0.639
Observations	5863	2935	2928		
F-statistic: 0.949		p-value: 0.504			
Panel C: Sample Composition and Attrition					
Present at Baseline (%)	81.63	79.80	83.47	-3.67	0.072
Present at Midline (%)	84.85	84.05	85.66	-1.61	0.270
Present at Endline (%)	77.64	79.15	76.13	3.02	0.235
Present in All Waves (%)	59.15	58.47	59.84	-1.37	0.637
F-statistic: 3.002		p-value: 0.035			

Notes: This table presents balance tests for school characteristics, student characteristics, and baseline outcome variables. Panel C shows sample composition and attrition patterns across waves. The last two columns show the difference between treatment and control means and the p-value from a t-test of equality of means. The F-statistic and corresponding p-value are from a joint test of significance of all variables in predicting treatment assignment within each panel. All academic and cognitive measures are standardized to have mean 50 and standard deviation 10 in the control group.

is an indicator for whether school s was assigned to the treatment group, Strata_s is a vector of stratification variables (indicators for tribal schools, girls' schools, and rural schools), $Y_{is}^{baseline}$ represents baseline controls for the same outcome when available, X_{is} is a vector of individual-level controls (age, socioeconomic status, computer skills, and parent occupation categories), and S_s represents school-level controls (building status, toilet facilities, wall, library, playground, water access, medical facilities, and accessibility ramps). Standard errors are clustered at the school level to account for within-school correlation in outcomes.

For outcomes that were not collected at baseline (problem-solving skills and grit), I omit the baseline control but include the full set of other controls and stratification variables.

For team-level outcomes, particularly the quality of innovation ideas collected at endline, I use a similar specification:

$$Y_{ts} = \alpha + \beta \text{ Treatment}_s + \gamma \text{ Girls}_s + \lambda \text{ Tribal}_s + \psi S_s + \phi X_{ts} + \varepsilon_{ts} \quad (2)$$

where Y_{ts} represents the outcome for team t in school s, and other variables are defined as above.

Given the multiple outcomes examined in this study, I address concerns about false positives by implementing two approaches to multiple hypothesis testing. First, I use the Romano-Wolf step-down procedure (Romano and Wolf, 2005) to adjust p-values across families of outcomes. Second, I implement the wild cluster bootstrap procedure (Cameron et al., 2008) to obtain more accurate p-values for my primary treatment effect estimates, given the cluster-randomized design.

5.4 Differential Attrition and Attendance as an Outcome

I work in a setting where student absenteeism is common due to various factors including family obligations, seasonal work, and health issues. The sample comprises 5,863 students across 80 schools. Attendance patterns varied across waves, with control schools showing higher participation at baseline (83.5% vs. 79.8%), similar rates at midline, and lower attendance at endline (76.1% vs. 79.1%).

This differential attendance could be interpreted in two ways. First, as a positive outcome of the intervention - the innovation training appears to have increased student engagement with school, leading to improved attendance. Second, as a potential threat to the internal validity of my treatment effect estimates on other outcomes, if the composition of students present at endline differs systematically between treatment and control schools.

To address this second concern, I conducted an extensive analysis of baseline characteristics among students present at endline. Among endline attendees, treatment and control groups remain well-balanced on all key baseline measures as shown in Table **??**. The joint F-test confirms no systematic differences between treatment and control students in the analyzed sample. I also examined whether baseline characteristics predict differential attrition by treatment status through interaction tests. The joint test of interactions between treatment and baseline academic and innovation measures yields a chi-squared of 0.25 ($p = 0.969$), demonstrating that the intervention did not fundamentally alter the relationship between student characteristics and attendance patterns. This

suggests that while attendance patterns differed between treatment and control schools, the composition of students remained comparable.

6. Results

This section presents the estimated treatment effects of the two-year innovation education program on students' outcomes. I organize the presentation by outcome category: innovation skills and higher-order cognitive skills, academic performance, and student behavior.

Table 2: Treatment Effects on Innovation Skills and Higher-Order Cognitive Skills (in Standard Deviations)

Outcome	Treatment Effect	WB p-value	MHT p-value	Observations
Panel A: Individual Innovation Measures				
Innovation Score	0.09^{**}	0.039	0.06	5,863
	(0.06)			
Exploitative Approach	0.13^{***}	0.008	0.01	5,863
	(0.04)			
Explorative Approach	0.07^{**}	0.041	0.04	5,863
	(0.04)			
Started With Exploration	0.14^{***}	0.000	0.01	5,863
	(0.05)			
Panel B: Innovation Team Outcomes				
Innovation Scale	0.24^{***}	0.010	0.02	1,189
	(0.09)			
Funding Amount (Std)	0.18^{**}	0.040	0.09	1,189
	(0.10)			
Panel C: Higher-Order Skills				
Fluid Intelligence	0.13^{**}	0.037	0.07	5,863
	(0.06)			
Emotional Intelligence	0.08	0.246	0.14	5,863
	(0.05)			
Problem Solving	0.19^{**}	0.020	0.04	5,863
	(0.07)			
Grit	-0.06	0.446	0.50	5,863
	(0.09)			
Panel D: Student Engagement with Innovation				
Liked Innovation Program	0.21^{***}	0.000	0.00	5,863
	(0.04)			
Daily Time on Innovation	0.22^{***}	0.010	0.01	5,863
	(0.07)			

Notes: Each cell reports the coefficient on the treatment indicator from a separate regression, expressed in standard deviation units of the control group. Values in parentheses are standard errors, clustered at the school level. WB p-value is the wild bootstrap p-value, calculated using 1000 replications. MHT p-value is the multiple hypothesis testing p-value using the Romano-Wolf procedure. Statistical significance is denoted by $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. All regressions include strata fixed effects, the outcome's baseline value when available, relevant domain-specific controls, demographic controls, and school characteristics. Innovation Team Outcomes are based on team-level assessments.

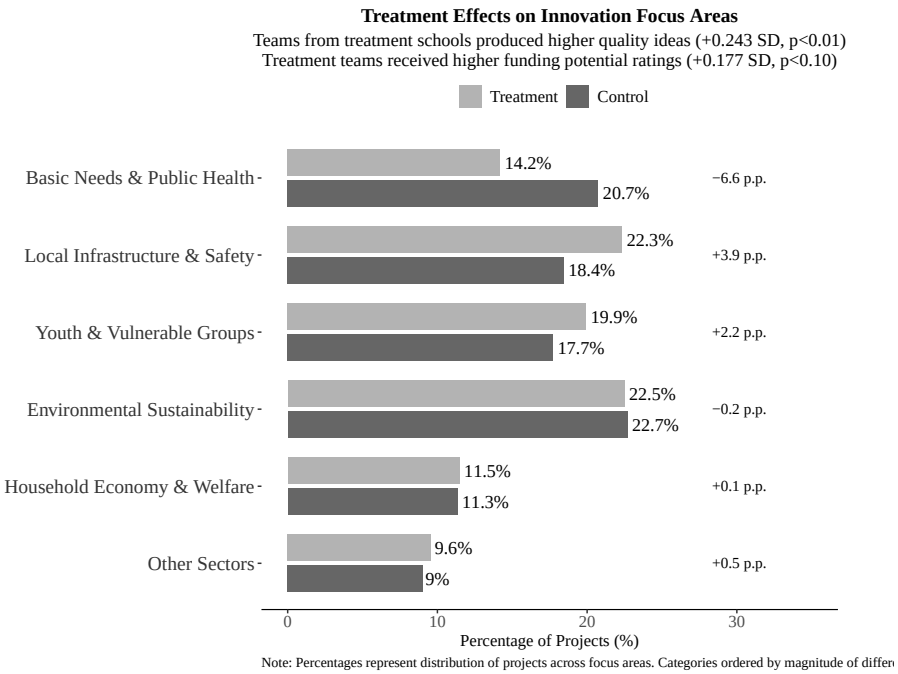


FIGURE 2: TREATMENT EFFECTS ON INNOVATION FOCUS AREAS

6.1 Treatment Effects on Innovation Skills

Table 2, Panel A presents treatment effects on individual innovation measures. In the multi-armed bandit task, treatment students demonstrated significantly greater skill in both key aspects of the innovation process. They scored 0.07 SD higher on explorative search (p=0.041), showing improved ability to test multiple options to discover potentially valuable alternatives. They also scored 0.13 SD higher on exploitative refinement (p=0.008), indicating stronger ability to capitalize on promising options once identified.

Importantly, treatment students were 0.14 SD more likely to begin with exploration rather than exploitation (p<0.001), reflecting the understanding that initial broad search typically precedes focused refinement in successful innovation processes. The combination of these improved decision-making strategies translated into tangible results, as evidenced by treatment students' 0.09 SD higher overall innovation scores (p=0.039). This indicates they were able to discover higher-value options in the bandit task, demonstrating that their improved exploration-exploitation strategies led to better outcomes rather than just different approaches.

As shown in Panel D, student engagement with innovation activities also increased in treatment schools, with higher self-reported engagement (0.21 SD, p<0.01) and more time spent on innovation-related activities (0.22 SD, p<0.01).

Teams from treatment schools produced ideas rated 0.24 SD higher on the innovation quality scale (p<0.01) and received funding potential ratings 0.18 SD higher (p<0.10)

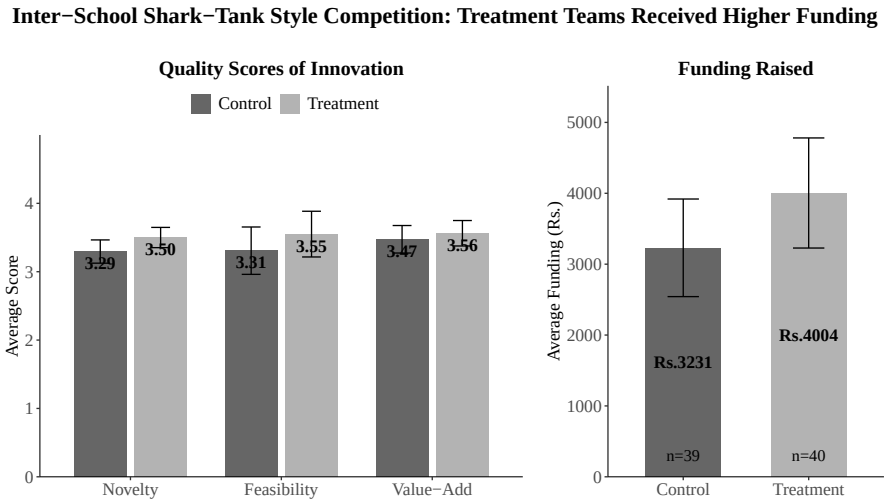


FIGURE 3: INTER-SCHOOL SHARK-TANK STYLE COMPETITION OUTCOMES

from expert evaluators.

Figure 2 presents the distribution of innovation domains addressed by students in treatment and control schools showing that the structured training did not channel students toward any single innovation domain, as evidenced by the diverse distribution of focus areas across both groups. Despite receiving identical training materials and methodologies, treatment students maintained diversity in their innovation focus. While some modest shifts in domain emphasis occurred, no single domain dominated the treatment group's attention. The relatively small magnitude of these shifts (all under 7 percentage points) demonstrates that the structured innovation program improved innovative quality without narrowing students' problem selection.

For the top 40 teams from each group that advanced to the inter-school competition, expert judges evaluated the innovation projects across multiple dimensions. As shown in Figure 3, treatment teams significantly outperformed control teams across all dimensions but since this is a much smaller sample size, I do not have statistical power to detect treatment effects. Treatment innovations received a mean funding recommendation of ₹771 more than control teams, representing a 24 p.p. higher than the control mean. The highest-funded treatment innovation was a health monitoring system for cardiac patients, while the highest-funded control innovation was an automatic farming machine. These findings suggest that students naturally gravitate toward addressing locally-relevant challenges when given opportunities to innovate, and that structured innovation training improves the quality of their solutions without constraining their problem selection.

6.2 Treatment Effects on Higher-Order Skills

Table 2, Panel C presents effects on higher-order cognitive and non-cognitive skills. Treatment students showed improvements in fluid intelligence (0.13 SD, p<0.05) and problem-solving ability (0.19 SD, p<0.05).

The program had smaller and statistically insignificant effects on emotional intelligence (0.08 SD, p>0.10) and grit (-0.06 SD, p>0.10). All of these findings remain consistent after adjusting for multiple hypothesis testing.

6.3 Treatment Effects on Academic Outcomes and Student Behavior

Table 3: Treatment Effects on Academic Outcomes and Student Behavior (in Standard Deviations)

Outcome	Treatment Effect	WB p-value	MHT p-value	Observations
Panel A: Academic Performance				
Math Score	0.07	0.29	0.69	5,863
	(0.06)			
Science Score	0.06	0.35	0.69	5,863
	(0.06)			
Panel B: Academic Engagement				
Liked Math	-0.02	0.70	0.69	5,863
	(0.05)			
Liked Science	-0.03	0.65	0.69	5,863
	(0.07)			
Daily Time on Homework	-0.05	0.54	0.69	5,863
	(0.08)			
Panel C: Student Behavior and Aspirations				
Attendance	0.08^{*}	0.10	0.26	5,863
	(0.05)			
Questions Asked	-0.05	0.55	0.54	5,863
	(0.07)			
Number of Friends	0.06	0.52	0.53	5,863
	(0.08)			
STEM Career Aspiration	-0.02	0.56	0.53	5,863
	(0.04)			
Teacher Career Aspiration	0.16^{***}	0.01	0.01	5,863
	(0.05)			

Notes: Each cell reports the coefficient on the treatment indicator from a separate regression, expressed in standard deviation units of the control group. Values in parentheses are standard errors, clustered at the school level. WB p-value is the wild bootstrap p-value, calculated using 1000 replications. MHT p-value is the multiple hypothesis testing p-value using the Romano-Wolf procedure. Statistical significance is denoted by $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. All regressions include strata fixed effects, the outcome's baseline value when available, relevant domain-specific controls,demographic controls, and school characteristics.

Table 3, Panels A and B show the program's effects on academic performance and engagement. Across the full sample, the innovation program had small positive but statistically insignificant effects on mathematics (0.07 SD, p=0.29) and science scores (0.06 SD, p=0.35).

The program had minimal impact on students' self-reported engagement with aca-

demic subjects. Treatment students showed small, statistically insignificant decreases in their reported liking of math (-0.02 SD, p=0.70) and science (-0.03 SD, p=0.65), and spent slightly less time on homework (-0.05 SD, p=0.54).

Table 3, Panel C presents effects on additional student outcomes. Treatment students showed marginally higher attendance rates (0.08 SD, p=0.10), and a highly significant increase in teacher career aspirations (0.16 SD, p=0.01). This may be due to peer-led model of the program. The program had no significant effect on number of questions asked in class (-0.05 SD, p=0.55), number of friends (0.06 SD, p=0.52), or STEM career aspirations (-0.02 SD, p=0.56).

These results suggest that the innovation curriculum significantly improved students' innovation skills and related higher-order skills. The increased attendance and teacher career aspirations may indicate that the program enhanced students' overall engagement with educational processes.

6.4 Heterogeneous Treatment Effects

This section examines how the innovation training program's impacts vary by students' baseline characteristics. I focus on three dimensions of heterogeneity: baseline math achievement, gender, and socioeconomic status.

6.4.1 Heterogeneity by Baseline Math Achievement

Table 4 presents treatment effects by baseline math achievement, showing clear differences in how the program affected students at different academic levels. The most notable pattern is a divergence in math performance: below-median students experienced a significant decline of 0.63 standard deviations (-0.63 SD) in math scores, equivalent to approximately 1.06 percentage points ($p < 0.05$), while above-median students showed significant improvement of 1.22 SD, corresponding to roughly 1.69 percentage points ($p < 0.05$).

Despite this divergent pattern in math, both groups showed significant gains in innovation skills, with innovation scores increasing by 1.05 SD (approximately 1.04 percentage points) for below-median students ($p < 0.10$) and 0.88 SD (about 1.17 percentage points) for above-median students ($p < 0.10$). Higher-order skill development also followed a consistent pattern favoring academically stronger students, with larger gains for above-median students in fluid intelligence (3.27 SD, approximately 3.99 percentage points, $p < 0.10$) compared to below-median students (1.20 SD, roughly 1.43 percentage points, not significant). Particularly noteworthy are the improvements in problem-solving abilities, with gains of 3.35 SD (about 3.25 percentage points) for below-median students and 4.36 SD (approximately 4.46 percentage points, $p < 0.05$) for above-median students.

Table 4: Heterogeneous Treatment Effects by Baseline Math Achievement

	Below Median Math	Above Median Math
Outcome	Treatment Effect	Treatment Effect
Panel A: Innovation Outcomes		
Innovation Score	1.05*	.882*
	(.6080)	(.625)
Exploitative Approach	2.33*	3.231**
	(1.198)	(1.147)
Explorative Approach	.894	.798
	(1.226)	(1.202)
Started With Exploration	.044**	.093***
	(.021)	(.027)
Panel B: Higher-Order Skills		
Fluid Intelligence	1.199	3.271*
	(1.523)	(1.535)
Emotional Intelligence	.333	1.321
	(1.081)	(.89)
Problem Solving	3.351	4.355**
	(2.105)	(1.731)
Grit	-.918	-1.984
	(2.883)	(3.31)
Panel C: Academic Outcomes		
Math Score	-.632**	1.215**
	(0.3)	(0.601)
Science Score	.084	.664
	(.784)	(.893)
Panel D: Selected Other Outcomes		
Attendance	2.658*	1.111
	(1.259)	(1.051)
STEM Career Aspiration	-.043	.004
	(.03)	(.034)
Teacher Career Aspiration	.02**	.018**
	(.008)	(.008)

Notes: Each row reports treatment effects from separate regressions. Below Median Math includes students scoring below the median (40%) on the baseline math assessment; Above Median Math includes those scoring at or above the median. Values in parentheses are standard errors, clustered at the school level. Statistical significance is denoted by $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. All regressions include strata fixed effects, the outcome's baseline value when available, demographic controls, and school characteristics.

6.4.2 Heterogeneity by Gender

Gender differences emerge in both innovation content and treatment impacts. Figure 4 shows that girls' teams produced higher quality innovations than boys' teams (+0.198 SD, p<0.05). The content of these innovations also differed, with girls focusing more on youth and vulnerable groups (+5.5 percentage points) and basic needs (+2.4 percentage points), while boys emphasized environmental sustainability (-10.5 percentage points) and infrastructure (-2.9 percentage points).

Table 5 reveals that treatment effects were generally larger for girls in innovation

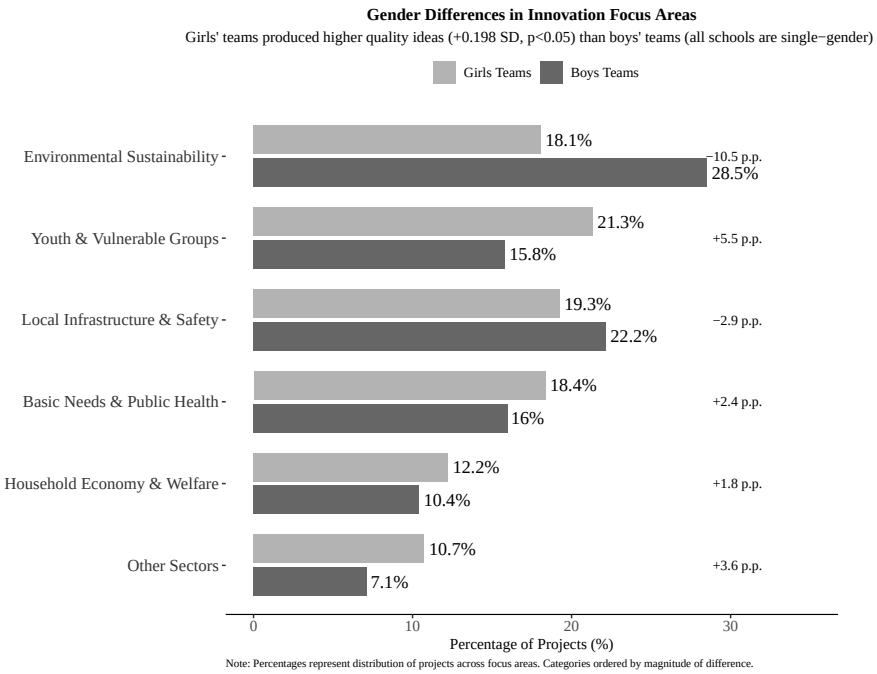


FIGURE 4: GENDER DIFFERENCES IN INNOVATION FOCUS AREAS

scores (+0.146 SD, equivalent to +1.97 percentage points, $p < 0.10$) and fluid intelligence (+0.196 SD, approximately +4.64 percentage points, $p < 0.10$). Boys, however, showed stronger gains in problem-solving (+0.306 SD, about +8.17 percentage points, $p < 0.05$). Girls in treatment schools also showed increased interest in teaching careers (+0.139 SD, equivalent to +3.4 percentage points, $p < 0.05$), with no significant effect observed in boys' schools. Additionally, boys' schools demonstrated improved attendance (+0.157 SD, $p < 0.05$) and modest gains in exploitative innovation approaches (+0.123 SD, $p < 0.10$), suggesting different responses to the program by gender.

These patterns suggest that innovation training engages boys and girls differently, with girls demonstrating stronger overall innovation performance and focusing more on social innovation domains, while boys show particular growth in problem-solving abilities and technical innovation areas.

6.4.3 Heterogeneity by Socioeconomic Status

Table 6 presents heterogeneous treatment effects by socioeconomic status (SES). Lower SES students experienced substantially stronger benefits from the innovation program in several key domains. Their innovation scores improved by 0.107 SD (approximately 1.87 percentage points, $p < 0.10$), compared to just 0.003 SD (about 0.63 percentage points) for higher SES students (not statistically significant).

Lower SES students also showed larger gains in fluid intelligence (0.117 SD, equivalent

Table 5: Heterogeneous Treatment Effects by School Gender Composition

	Boys' Schools	Girls' Schools
Outcome	Treatment Effect	Treatment Effect
Panel A: Innovation Outcomes		
Innovation Score	.029	.146*
	(.073)	(.06)
Exploitative Approach	.123*	.058
	(.06)	(.066)
Explorative Approach	.054	-.009
	(.058)	(.066)
Started With Exploration	.093	.231**
	(.054)	(.064)
Panel B: Higher-Order Skills		
Fluid Intelligence	.094	.196*
	(.069)	(.087)
Emotional Intelligence	.059	.038
	(.074)	(.065)
Problem Solving	.306**	.062
	(.105)	(.079)
Grit	.014	-.049
	(.099)	(.128)
Panel C: Academic Outcomes		
Math Score	-.007	.078
	(.098)	(.089)
Science Score	.06	.088
	(.073)	(.101)
Panel D: Selected Other Outcomes		
Attendance	.157**	.039
	(.065)	(.064)
STEM Career Aspiration	.02	-.061
	(.038)	(.086)
Teacher Career Aspiration	.068	.139**
	(.07)	(.049)

Notes: Each row reports treatment effects from separate regressions, with all outcomes standardized by the control group standard deviation within each school type. Values in parentheses are standard errors, clustered at the school level. Statistical significance is denoted by $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. All regressions include strata fixed effects, the outcome's baseline value when available, demographic controls, and school characteristics.

to 3.10 percentage points, $p < 0.10$) and problem-solving skills (0.168 SD, about 4.19 percentage points, $p < 0.05$) than their higher SES peers. While math performance showed no significant effects in either group, the direction of effects differed slightly: positive for lower SES students (0.035 SD, +0.50 percentage points) and negative for higher SES students (-0.005 SD, -0.31 percentage points).

Both groups showed significant improvements in certain innovation practices, with started exploration displaying the largest effects (0.126 SD for lower SES, $p < 0.05$; 0.210 SD for higher SES, $p < 0.01$). Notably, lower SES students were more likely to aspire to teaching careers after the intervention (0.157 SD increase, approximately 3.84 percentage

Table 6: Heterogeneous Treatment Effects by Socioeconomic Status

	Lower SES	Higher SES
Outcome	Treatment Effect	Treatment Effect
Panel A: Innovation Outcomes		
Innovation Score	.107*	.003
	(.058)	(.054)
Exploitative Approach	.113*	.099*
	(.051)	(.054)
Explorative Approach	.008	.067
	(.047)	(.05)
Started With Exploration	.126**	.21***
	(.047)	(.058)
Panel B: Higher-Order Skills		
Fluid Intelligence	.117*	.07
	(.067)	(.07)
Emotional Intelligence	.058	.031
	(.052)	(.065)
Problem Solving	.168**	.108
	(.071)	(.08)
Grit	-.044	-.067
	(.084)	(.102)
Panel C: Academic Outcomes		
Math Score	.035	-.005
	(.073)	(.067)
Science Score	.031	.084
	(.067)	(.07)
Panel D: Selected Other Outcomes		
Attendance	.089	.105
	(.053)	(.059)
STEM Career Aspiration	-.035	-.055
	(.062)	(.065)
Teacher Career Aspiration	.157**	.063
	(.057)	(.058)

Notes: Each row reports treatment effects from separate regressions, with all outcomes standardized by the control group standard deviation within each SES group. Lower SES includes students from low and lower-middle SES backgrounds; Higher SES includes those from middle and upper SES households. Values in parentheses are standard errors, clustered at the school level. Statistical significance is denoted by $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. All regressions include strata fixed effects, the outcome's baseline value when available, demographic controls, and school characteristics.

points, $p < 0.05$), with no significant change among higher SES students.

Career aspirations also showed a heterogeneous pattern, with teacher career aspirations increasing significantly among lower SES students (2.8 percentage points, p<0.05) but not among higher SES students. These results suggest that structured innovation training may be particularly beneficial for students from disadvantaged backgrounds, providing them opportunities to develop skills that might otherwise be less accessible in their educational environment.

6.5 Robustness and Sensitivity Analysis

I conducted several tests to assess the robustness of treatment effect estimates to potential attrition bias. Table **??** presents treatment effects estimated under three different analytical approaches. For innovation scores, the primary outcome of interest, the estimated treatment effect remains consistent: 1.23 points ($p = 0.119$) using the endline sample, 1.25 points ($p = 0.114$) with inverse probability weighting to adjust for predicted attrition, and 0.92 points ($p = 0.318$) when restricting to the balanced panel of students present at all three waves. While statistical significance varies across specifications, the coefficient magnitudes remain stable, suggesting that differential attrition does not substantially bias the main treatment effect estimates.

To further examine potential selection bias, Table **??** shows that treatment and control students present at endline maintain balance on all measured baseline characteristics. The joint F-test for baseline balance ($F(13, 79) = 1.06$, $p = 0.408$) confirms no systematic differences between treatment and control students in the analyzed sample. Additionally, the test for differential selection ($\chi^2(3) = 0.25$, $p = 0.969$) indicates that baseline academic and innovation measures do not predict attrition differently across treatment and control groups, providing further evidence against selection bias.

The main results in Tables 2 and 3 employ both wild cluster bootstrap p-values and multiple hypothesis testing adjustments to ensure robust statistical inference. The wild cluster bootstrap approach addresses concerns about inference with a moderate number of clusters (80 schools), while the Romano-Wolf step-down procedure controls for the family-wise error rate when testing multiple outcomes within each domain. The consistency of findings across both conventional clustering and these more conservative inferential approaches provides strong evidence that the key results are not statistical artifacts. Notably, the improvements in innovation skills, problem-solving, and fluid intelligence remain statistically significant even after these adjustments, further strengthening confidence in the primary findings.

These analyses support the internal validity of the treatment effects reported in the main results. The consistency of estimates across different specifications and inferential approaches, alongside the absence of differential selection on observables, suggests that the findings are not driven by compositional changes in the sample over time or by statistical artifacts related to multiple hypothesis testing or cluster structure.

7. Discussion

The innovation training program produced heterogeneous effects that reveal tensions within India's educational system. Both high and low-performing students developed

innovation skills, problem-solving abilities, and fluid intelligence, but experienced different academic outcomes. These findings can be understood through a skill production framework where innovation and academic skills draw from a common pool of cognitive resources with different productivity parameters that vary by student type. For students with above-median baseline math performance, the innovation program produced complementary outcomes. These students showed gains in both innovation skills and mathematics performance, consistent with Cunha and Heckman (2007)'s finding that foundational skills facilitate transfer across learning domains. For these students, innovation training appears to generate spillovers, affecting the productivity parameter for both skill types.

In contrast, below-median students improved in innovation and problem-solving but showed reduced math performance and motivation. This suggests a reallocation of cognitive resources toward innovation skills where their comparative productivity improved through training—aligning with Becker (1965)'s time allocation framework. This reallocation may represent specialization if innovation skills offer these students higher returns than marginal improvements in math performance.

These findings indicate that academic and higher-order skills do not necessarily develop in tandem. The innovation program offered struggling students an alternative pathway to develop higher-order thinking skills while potentially widening the academic achievement gap. This presents policymakers with a choice: should education policies prioritize convergence in standardized test metrics, or should they facilitate skill diversification that may match students' comparative advantages and labor market needs?

This pattern also revealed a contradiction within India's educational policy framework—a system advocating for higher-order skill development while evaluating students through memorization-focused assessments. India's National Education Policy 2020 aims to develop "creative and critical thinking" and "problem-solving abilities," while initiatives like Atal Tinkering Labs invest in innovation infrastructure. Yet classroom practices feature rote learning, with 90% of teacher questions requiring recall (Bhattacharjea et al., 2011). Students face examinations that reward memorization rather than applied reasoning.

This contradiction reflects the dual purposes of education in India—as both a human capital development system and a sorting mechanism for limited higher education opportunities. As Muralidharan (2024) argues, when education functions to identify "high-performing" students for advanced opportunities, assessment practices prioritize discrimination between students over skill development.

The separation between academic and applied knowledge affects marginalized students who lack home resources to bridge this gap independently. These results align with Banerjee et al. (2025)'s finding that students can solve calculation problems in real-world

contexts yet struggle with identical mathematical concepts in academic formats requiring memorized procedures. The innovation program showed that students categorized as "low-performing" in traditional metrics develop higher-order thinking when engaged with applied, contextually relevant content.

The pattern of skill improvements observed in this study indicates mechanisms through which innovation education may influence cognitive development. The program improved three domains: innovation skills (0.12-0.24 SD), problem-solving abilities (0.19 SD), and fluid intelligence (0.13 SD). However, it showed limited impact on grit and emotional intelligence. This pattern suggests the intervention enhanced domain-general abstract reasoning capabilities rather than non-cognitive traits—a distinction for understanding which dimensions of human capital respond to educational interventions during adolescence.

An alternative explanation for these patterns could be measurement error or insufficient intervention intensity for non-cognitive traits. However, both grit and emotional intelligence were measured using behavioral tasks rather than self-reports, reducing social desirability concerns. The consistent pattern across multiple cognitive measures but not non-cognitive ones indicates that the intervention targets cognitive flexibility rather than personality traits.

The improvement in fluid intelligence merits attention given research indicating such abilities maintain stability. The intervention's iterative project cycles, requiring students to recognize patterns across different technologies and adapt them to new contexts, may have activated abstract reasoning pathways. This aligns with research on neuroplasticity suggesting educational interventions can enhance fluid intelligence during adolescence, a period of prefrontal cortex development (Klingberg, 2010; Diamond and Lee, 2011). The magnitude of fluid intelligence improvements (0.13 SD) corresponds to effects found in targeted cognitive training programs (Au et al., 2015), suggesting innovation education could serve as a contextually embedded approach to developing this capability.

The parallel improvements in innovation and problem-solving indicate these abilities may develop through related but distinct cognitive mechanisms. Both require students to identify patterns and generate solutions, but they differ in constraint definition. Problem-solving tasks (like the Sudoku puzzles used in assessment) present defined constraints with clear solution paths, while innovation challenges require students to operate under uncertainty with ambiguous constraints. The program's improvement of both abilities indicates it developed students' capacity to work within both defined and ambiguous problem spaces.

I also show demographic characteristics influence content of students' innovations. The innovation ideas generated by students focused on challenges relevant to their envi-

ronments and communities. Both treatment and control students directed attention to agriculture, water management, environmental sustainability, and educational access—areas reflecting challenges faced by communities in rural Telangana. Girls' teams produced higher quality innovations than boys' teams (+0.198 SD) and focused more on youth, vulnerable groups, and basic needs, while boys emphasized environmental sustainability and infrastructure. This pattern aligns with Koning et al. (2021)'s finding that innovator demographics influence problem selection, extending this principle to educational settings.

Treatment students produced higher quality ideas (0.24 SD) and secured more funding in the shark-tank style competition, and the distribution of innovation domains remained diverse. The structured innovation training improved quality without channeling students toward any single domain, with shifts in focus areas remaining under 7 percentage points. This indicates the program enhanced innovative capacity without constraining problem selection, a distinction from more prescriptive educational approaches.

I developed and validated new measurement tools for assessing economically valuable higher-order skills in resource-constrained environments, contributing methodologically to educational research. These instruments address a gap in educational research in developing countries, where technology infrastructure limitations exist. The multi-armed bandit task I adapted from Ederer and Manso (2013a) provides a behavioral measure of innovation that captures both exploratory tendencies and optimization abilities. By contextualizing this task within familiar local businesses, I ensured the instrument maintains theoretical connections while ensuring cultural relevance.

The Innovation-S scale I developed through consultation with practicing innovators provides a standardized rubric for assessing idea quality with predictive validity for real-world outcomes. The scale's ability to predict funding decisions in the shark-tank style competition indicates its value as an assessment tool. These instruments I created advance beyond traditional self-reported measures of higher-order skills, which face limitations including social desirability bias and reference group effects.

8. Cost-Effectiveness and Policy Implications

8.1 Cost-Effectiveness

The innovation training program by Inqui-lab costs approximately $40 per student annually, primarily for materials, peer facilitator training, and monitoring. At this cost, the program generated improvements of 0.12–0.24 standard deviations in innovation skills, 0.15 standard deviations in problem-solving, and 0.14 standard deviations in fluid intelligence. These effect sizes are comparable to or larger than many educational interventions

in developing countries (Evans and Yuan, 2022), while the cost is substantially lower than many alternatives.

For comparison, computer-aided learning programs that have shown similar effect sizes in India cost approximately $100–$200 per student (Muralidharan et al., 2019a), while remedial education programs cost around $67 per student for comparable gains (Banerjee et al., 2007). At $40 per student, the innovation program achieves a cost-effectiveness ratio of approximately $182 per standard deviation improvement in innovation skills ($40 ÷ 0.22 SD), comparing favorably to other successful educational interventions in similar contexts.

8.2 Scaling and Policy Implications

Implementing this program for all Grade 8 students in Telangana would cost approximately $16 million annually, representing 0.6% of the state's education budget. This is comparable to existing investments in innovation education through programs like Atal Tinkering Labs ($240 million nationally, NITI Aayog 2021b) and Delhi's Entrepreneurship Mindset Curriculum ($27 million annually, Delhi Government 2019).

A back-of-the-envelope calculation can illustrate the potential long-run returns to such an investment. The intervention improved both the innovation skills of high-performing students (potential future inventors) and the problem-solving capabilities of students across the distribution. If these skills persist into adulthood, they could address documented productivity gaps in India's economy through two channels:

First, for the approximately 88% of workers likely to enter the small enterprise sector (ILO, 2018), enhanced problem-solving and innovation skills could help address inefficiencies similar to those targeted by other interventions in developing economies. Studies have documented productivity gains of 11–17% from improved management practices (Bloom et al., 2013), 8–10% from better information utilization (Jensen and Miller, 2018), 12% from enhanced soft skills (Adhvaryu et al., 2018), and 20–35% from technological adaptation (Atkin et al., 2017). If the skills developed through this intervention enabled even modest improvements in these domains—say 5–10% increased productivity for 20% of future workers—this would represent a meaningful economic impact.

Second, for students with high academic performance who may pursue STEM careers, improved innovation skills could enhance India's formal innovation output. Jones (2022) estimates that developing countries could quadruple innovation output through improved talent allocation and education. If the innovation training improved the effectiveness of even a small fraction of future STEM graduates, the aggregate effects could be substantial given the outsized impact of innovation on economic growth.

These calculations are necessarily speculative and depend on the persistence of skill

improvements and successful transition into the workforce. However, they illustrate that relatively modest investments in innovation skill development during adolescence could plausibly generate significant returns, particularly in a rapidly developing economy where both small enterprise productivity and formal innovation are important growth drivers.

9. Conclusion

This study provides experimental evidence that innovation skills can be systematically taught to disadvantaged youth in a developing country context. Through a large-scale randomized controlled trial in 80 residential schools in India, I demonstrate that a structured innovation training program improved innovative ability (0.12-0.24 SD), problem-solving skills (0.19 SD), and fluid intelligence (0.13 SD) among students from constitutionally protected marginalized communities.

The heterogeneous treatment effects observed in this study highlight important theoretical implications for understanding skill development. For high-performing students, innovation training generated complementary gains across both innovation and academic domains, consistent with theories of positive skill spillovers. For struggling students, the program led to improved higher-order thinking skills but reduced mathematics performance, suggesting a reallocation of cognitive resources toward domains of comparative advantage. These differential responses show that skill development is neither uniform nor unidimensional, challenging simplified models of human capital formation.

My findings also document how innovation education influences the content and focus of innovation itself. Students' innovations addressed challenges directly relevant to their socioeconomic context, and demographic characteristics—particularly gender—shaped which problems received attention. This pattern extends recent economic research on innovator demographics and problem selection into educational settings, indicating that diversifying who innovates may fundamentally alter innovation priorities even at the educational stage.

The methodological tools developed for this study advance the measurement of economically valuable higher-order skills in resource-constrained environments. By providing behaviorally-based assessments that function effectively without sophisticated technology infrastructure, these instruments contribute to the broader challenge of measuring non-cognitive skills in developing countries.

As detailed in my cost-effectiveness analysis, the program's impacts were achieved at a lower cost than many comparable interventions, with potential long-term economic returns through multiple channels. The intervention's impact on innovation quality was confirmed not only through standardized assessments but also through market-based

validation, with treated teams securing significantly more funding in a competitive environment judged by innovation experts.

The tensions observed between developing higher-order skills and traditional academic metrics reflect fundamental contradictions within educational systems designed primarily for sorting rather than skill development. Successfully developing innovation capabilities across the student distribution will require aligning assessment practices with stated educational goals and recognizing the diversity of skills that contribute to economic productivity.

Further research is needed to track the long-term persistence of these skill improvements and their translation into economic outcomes, to identify the most effective components of innovation training, and to develop approaches that can simultaneously enhance both traditional academic and higher-order skills for all students. Such research would further clarify how educational systems can more effectively develop the range of capabilities needed in rapidly evolving economies—particularly for historically marginalized students who represent a substantial reserve of untapped innovative potential.

References

Acemoglu, Daron and David Autor (2011). "Skills, tasks and technologies: Implications for employment and earnings". *Handbook of labor economics* 4, pp. 1043–1171.

Adhvaryu, Achyuta, Namrata Kala, and Anant Nyshadham (2018). "The skills to pay the bills: Returns to on-the-job soft skills training". *NBER Working Paper No. 24313*.

Aghion, Philippe, Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen (2017a). "The social origins of inventors". *NBER Working Paper No. 24110*.

— (2017b). *The social origins of inventors*. Tech. rep. National Bureau of Economic Research.

AISHE (2020). *All India Survey on Higher Education 2019-20*. Tech. rep. New Delhi: Department of Higher Education, Government of India.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi (2018). "Dancing with the Stars: Innovation Through Interactions". *NBER Working Paper* 24466.

Alan, Sule, Teodora Boneva, and Seda Ertac (2019a). "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit". *The Quarterly Journal of Economics* 134.3, pp. 1121–1162.

— (2019b). "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit". *Quarterly Journal of Economics* 134.3, pp. 1121–1162.

Alan, Sule and Seda Ertac (2018). "Fostering patience in the classroom: Results from randomized educational intervention". *Journal of Political Economy* 126.5, pp. 1865–1911.

Alan, Sule and Elif Kubilay (2025). "Empowering adolescents to transform schools: Lessons from a behavioral targeting". *American Economic Review* 115.2, pp. 365–407.

Alan, Sule and Ipek Mumcu (2024). "Nurturing childhood curiosity to enhance learning: Evidence from a randomized pedagogical intervention". *American Economic Review* 114.4, pp. 1173–1210.

AMS Public Relations Society (2022). *School Innovation Challenge Bootcamp 2022*. Tech. rep. AMS Public Relations Society.

ASER (2022). *Annual Status of Education Report (Rural) 2022*. Tech. rep. New Delhi: Pratham Education Foundation.

Asher, Sam, Paul Novosad, and Charlie Rafkin (2022). "Intergenerational Mobility in India: Estimates from New Methods and Administrative Data". *American Economic Review* 112.8, pp. 2456–2486.

Ashraf, Nava, Abhijit Banerjee, and Vesall Nourani (2020a). "Learning to teach by learning to learn". *Working paper*.

Ashraf, Nava, Natalie Bau, Corinne Low, and Kathleen McGinn (2020b). "Negotiating a better future: How interpersonal skills facilitate intergenerational investment". *The Quarterly Journal of Economics* 135.2, pp. 1095–1151.

Atkin, David, Amit K. Khandelwal, and Adam Osman (2017). "Exporting and firm performance: Evidence from a randomized experiment". *The Quarterly Journal of Economics* 132.2, pp. 551–615.

Au, Jacky, Ellen Sheehan, Nancy Tsai, Greg J Duncan, Martin Buschkuehl, and Susanne M Jaeggi (2015). "Improving fluid intelligence with training on working memory: a meta-analysis". *Psychonomic Bulletin & Review* 22.2, pp. 366–377.

Bandiera, Oriana, Stephen Hansen, Andrea Prat, and Raffaella Sadun (2021). "Autonomy and incentives: Evidence from a field experiment in Bangladesh". *Economic Journal* 131.639, pp. 2795–2830.

Bando, Rosangela, Emma Näslund-Hadley, and Paul Gertler (2019). "Effect of inquiry and problem based pedagogy on learning: Evidence from 10 field experiments in four countries". *NBER Working Paper No. 26280*.

Banerjee, Abhijit V., Suman Bhattacharjee, Raghabendra Chattopadhyay, Esther Duflo, Alejandro J. Ganimian, Kalpana Rajah, and Elizabeth S. Spelke (2025). "Children's arithmetic skills do not transfer between applied and academic mathematics". *Nature*. forthcoming.

Banerjee, Abhijit V., Shawn Cole, Esther Duflo, and Leigh Linden (2007). "Remedying education: Evidence from two randomized experiments in India". *The Quarterly Journal of Economics* 122.3, pp. 1235–1264.

Baron-Cohen, Simon, Sally Wheelwright, Jacqueline Hill, Yogini Raste, and Ian Plumb (2001). "The reading the mind in the eyes test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism". *Journal of Child Psychology and Psychiatry* 42.2, pp. 241–251.

Becker, Gary S. (1965). "A Theory of the Allocation of Time". *The Economic Journal* 75.299, pp. 493–517.

Bell, Alexander, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen (2019a). "Who becomes an inventor in America? The importance of exposure to innovation". *The Quarterly Journal of Economics* 134.2, pp. 647–713.

— (2019b). "Who becomes an inventor in America? The importance of exposure to innovation". *Quarterly Journal of Economics* 134.2, pp. 647–713.

Bettinger, Eric, Stein Ludvigsen, Mari Rege, Ingeborg F. Solli, and David Yeager (2018). "Increasing perseverance in math: Evidence from a field experiment in Norway". *Journal of Economic Behavior & Organization* 146, pp. 1–15.

Bhattacharjea, Suman, Wilima Wadhwa, and Rukmini Banerji (2011). "Inside Primary Schools: A Study of Teaching and Learning in Rural India".

Biasi, Barbara and Song Ma (2022). "The education-innovation gap". *NBER Working Paper No. 29853*.

Blikstein, Paulo, José Armando Valente, and Susana Moura (2022). "School Makerspaces in Developing Countries: Outcomes and Implementation Challenges". *International Journal of Child-Computer Interaction* 31, pp. 100–112.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts (2013). "Does management matter? Evidence from India". *The Quarterly Journal of Economics* 128.1, pp. 1–51.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller (2008). "Bootstrap-based improvements for inference with clustered errors". *The Review of Economics and Statistics* 90.3, pp. 414–427.

Carpenter, Patricia A, Marcel Adam Just, and Peter Shell (1990). "What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test". *Psychological Review* 97.3, pp. 404–431.

Chioda, Laura, David Contreras-Loya, Paul Gertler, and Dana Carney (2021). "Making entrepreneurs: Returns to training youth in hard versus soft business skills". *NBER Working Paper No. 28845*.

Cunha, Flavio and James Heckman (2007). "The technology of skill formation". *American Economic Review* 97.2, pp. 31–47.

Cunha, Flavio and James J. Heckman (2008). "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation". *Journal of Human Resources* 43.4, pp. 738–782.

Daw, Nathaniel D, John P O'doherty, Peter Dayan, Ben Seymour, and Raymond J Dolan (2006). "Cortical substrates for exploratory decisions in humans". *Nature* 441.7095, pp. 876–879.

Delhi Government (2019). *Delhi's Education Revolution: Annual Report 2019-20*. Tech. rep. New Delhi: Directorate of Education, Government of NCT of Delhi.

Deming, David J. (2022). "Four Facts about Human Capital". *Journal of Economic Literature* 60.3, pp. 839–866.

Diamond, Adele and Kathleen Lee (2011). "Interventions shown to aid executive function development in children 4 to 12 years old". *Science* 333.6045, pp. 959–964.

Djaker, Sharnic, Alejandro J Ganimian, and Shwetlena Sabarwal (2024). "Out of sight, out of mind? The gap between students' test performance and teachers' estimations in India and Bangladesh". *Economics of Education Review* 102, p. 102575.

Dong, Nianbo and Rebecca Maynard (2013). "PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies". *Journal of Research on Educational Effectiveness* 6.1, pp. 24–67.

Duckworth, Angela L, Christopher Peterson, Michael D Matthews, and Dennis R Kelly (2007). "Grit: perseverance and passion for long-term goals". *Journal of Personality and Social Psychology* 92.6, pp. 1087–1101.

Ederer, Florian and Gustavo Manso (2013a). "Is pay for performance detrimental to innovation?" *Management Science* 59.7, pp. 1496–1513.
— (2013b). "Is pay for performance detrimental to innovation?" *Management Science* 59.7, pp. 1496–1513.
Einiö, Elias, Josh Feng, and Xavier Jaravel (2023). "Social push and the direction of innovation". *The Quarterly Journal of Economics*. forthcoming.
Evans, David K. and Fei Yuan (2022). "Education in developing countries: What policies and programmes affect learning and time in school?" *Review of Education* 10.1, e3354.
Galton, Emma (2019). "Grit and entrepreneurial performance: A meta-analytic investigation". *Journal of Business Venturing Insights* 11, e00124.
Government of Telangana (2019). "Telangana Innovation Programs".
— (2022). *Telangana Socioeconomic Outlook 2022*. Tech. rep. Hyderabad: Government of Telangana.
Hedges, Larry V. and E. C. Hedberg (2007). "Intraclass Correlation Values for Planning Group-Randomized Trials in Education". *Educational Evaluation and Policy Analysis* 29.1, pp. 60–87.
Herz, Holger, Daniel Schunk, and Christian Zehnder (2014). "How do people learn to allocate resources? Comparing two learning theories". *Experimental Economics* 17.1, pp. 1–31.
Hill, Carolyn J., Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey (2008). "Empirical Benchmarks for Interpreting Effect Sizes in Research". *Child Development Perspectives* 2.3, pp. 172–177.
Hnatkovska, Viktoria, Amartya Lahiri, and Sourabh B. Paul (2013). "Breaking the Caste Barrier: Intergenerational Mobility in India". *Journal of Human Resources* 48.2, pp. 435–473.
Hoxby, Caroline (2000). "Peer effects in the classroom: Learning from gender and race variation". *The Quarterly Journal of Economics* 115.3, pp. 1239–1285.
ILO (2018). *Women and men in the informal economy: A statistical picture*. Report. Geneva: ILO.
India Skills Report (2022). *India Skills Report*. Tech. rep. Gurugram: Wheebox.
India's Ministry of Education (2020). "National Education Policy 2020".
Jachimowicz, Jon M, Andreas Wihler, Erica R Bailey, and Adam D Galinsky (2018). "Why grit requires perseverance and passion to positively predict performance". *Proceedings of the National Academy of Sciences* 115.40, pp. 9980–9985.
Jaravel, Xavier, Neviana Petkova, and Alex Bell (2018). "Competition and innovation: The breakup of IG Farben". *Working Paper*.
Jauk, Emanuel, Mathias Benedek, Beate Dunst, and Aljoscha C Neubauer (2013). "The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection". *Intelligence* 41.4, pp. 212–221.
Jensen, Robert and Nolan H. Miller (2018). "Market integration, demand, and the growth of firms: Evidence from a natural experiment in India". *American Economic Review* 108.12, pp. 3583–3625.
Jones, Benjamin F (2009). "The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?" *The Review of Economic Studies* 76.1, pp. 283–317.

Jones, Charles I. (2022). "The past and future of economic growth: A semi-endogenous perspective". *Annual Review of Economics* 14, pp. 125–152.

Klingberg, Torkel (2010). "Training and plasticity of working memory". *Trends in Cognitive Sciences* 14.7, pp. 317–324.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson (2021). "Who do we invent for? Patents by women focus more on women's health, but few women get to invent". *Science* 372.6548, pp. 1345–1348.

Kushagra, A. (2023). "Shark Tank India Clones: How Reality TV Shows Are Fuelling Interest In Startup Ecosystem". *Inc42*.

Laureiro-Martínez, Daniella, Stefano Brusoni, Nicola Canessa, and Maurizio Zollo (2015). "Neural foundations of Reasoning in Humans by Managers Exploring versus Exploiting". *Scientific reports* 5.1, pp. 1–11.

March, James G. (1991a). "Exploration and exploitation in organizational learning". *Organization Science* 2.1, pp. 71–87.

March, James G (1991b). "Exploration and exploitation in organizational learning". *Organization Science* 2.1, pp. 71–87.

McKenzie, David (2012). "Beyond Baseline and Follow-up: The Case for More T in Experiments". *Journal of Development Economics* 99.2, pp. 210–221.

Ministry of Education, Thailand (2018). *Coding for Kids Initiative: Implementation Guidelines*. Tech. rep. Government of Thailand.

Ministry of ICT and Innovation, Rwanda (2019). *National Coding Curriculum Framework*. Tech. rep. Government of Rwanda.

Mooradian, Todd, Kurt Matzler, Borislav Uzelac, and Florian Bauer (2016). "Perspiration and inspiration: Grit and innovativeness as antecedents of entrepreneurial success". *Journal of Economic Psychology* 56, pp. 232–243.

Muralidharan, Karthik (2024). *Accelerating India's Development: A State-led Roadmap for Effective Governance*. Penguin Viking.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian (2019a). "Experimental evidence on the effects of technology-aided instruction in schools". *Journal of Development Economics* 104, p. 104919.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J Ganimian (2019b). "Personalized education: Experimental evidence from a computer-assisted learning program". *American Economic Review* 109.4, pp. 1426–1460.

Newell, Allen and Herbert A Simon (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

NITI Aayog (2020). *Atal Innovation Mission: Fostering Innovation and Entrepreneurship in Schools*. Tech. rep. Government of India.

— (2021a). "Atal Innovation Mission". Retrieved from https://aim.gov.in/.

— (2021b). *Atal Innovation Mission Annual Report 2020-21*. Tech. rep. New Delhi: NITI Aayog, Government of India.

OECD (2015). *Schools for 21st-Century Learners: Strong Leaders, Confident Teachers, Innovative Approaches*. Tech. rep. Organisation for Economic Co-operation and Development.

— (2018). *Oslo Manual 2018: Guidelines for collecting, reporting and using data on innovation*. Manual. Paris: OECD.

PLFS (2022). *Periodic Labour Force Survey (PLFS) - Annual Report (July 2021 - June 2022)*. Tech. rep. New Delhi: National Statistical Office, Government of India.

Raven, John (2000). "The Raven's progressive matrices: change and stability over culture and time". *Cognitive psychology* 41.1, pp. 1–48.

Raven, John C (1938). "Progressive matrices: A perceptual test of intelligence". *London: HK Lewis*.

Razzouk, Rim and Valerie Shute (2012). "What Is Design Thinking and Why Is It Important?" *Review of Educational Research* 82.3, pp. 330–348.

Romano, Joseph P and Michael Wolf (2005). "Stepwise multiple testing as formalized data snooping". *Econometrica* 73.4, pp. 1237–1282.

Russell, Stuart J. and Peter Norvig (2010). *Artificial Intelligence: A Modern Approach*. 3rd. Upper Saddle River, New Jersey: Prentice Hall.

Sacerdote, Bruce (2014). "Experimental and quasi-experimental analysis of peer effects: two steps forward?" *Annual Review of Economics* 6.1, pp. 253–272.

Simon, Herbert A (1973). "The structure of ill structured problems". *Artificial Intelligence* 4.3-4, pp. 181–201.

Soni, J. and K. Sharma (2022). "The Shark Tank India effect: 62% surge in entrepreneurship, startup pitches among school & college students". *The Economic Times*.

Telangana State Innovation Cell (2021). *UNICEF-TSIC School Innovation Challenge*. Tech. rep. Government of Telangana.

Toivanen, Otto and Lotta Väänänen (2016). "Education and invention". *Review of Economics and Statistics* 98.2, pp. 382–396.

UNESCO (2017). *Cracking the Code: Girls' and Women's Education in Science, Technology, Engineering and Mathematics (STEM)*. Tech. rep. United Nations Educational, Scientific and Cultural Organization.

— (2018). *UIS Fact Sheet: How much does your country invest in R&D?* Fact Sheet. UNESCO.

Vellante, Marcello, Simon Baron-Cohen, Marianna Melis, Matteo Marrone, Donatella Rita Petretto, Carmelo Masala, and Antonio Preti (2013). "The "Reading the Mind in the Eyes" test: Systematic review of psychometric properties and a validation study in Italy". *Cognitive Neuropsychiatry* 18.4, pp. 326–354.

Wadhwa, P. (2022). "How Shark Tank India is creating a new generation of entrepreneurs". *Business Standard*.

WIPO (2022). *World Intellectual Property Indicators 2022*. Report. Geneva: WIPO.

World Bank (2016). *Science, Technology and Skills for Africa's Development*. Tech. rep. World Bank Africa Region.

Wuchty, Stefan, Benjamin F Jones, and Brian Uzzi (2007). "The increasing dominance of teams in production of knowledge". *Science* 316.5827, pp. 1036–1039.

Yato, Takayuki and Takahiro Seta (2003). "Complexity and completeness of finding another solution and its application to puzzles". *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*. Vol. 86. 5, pp. 1052–1060.

Yeager, David S., Paul Hanselman, Gregory M. Walton, Jared S. Murray, Robert Crosnoe, Chandra Muller, et al. (2019). "A national experiment reveals where a growth mindset improves achievement". *Nature* 573.7774, pp. 364–369.

Zárate, Román Andrés (2022). "Social and academic network effects on learning: Evidence from a field experiment". *American Economic Journal: Applied Economics* 14.3, pp. 256–287.

10. Additional Figures and Tables

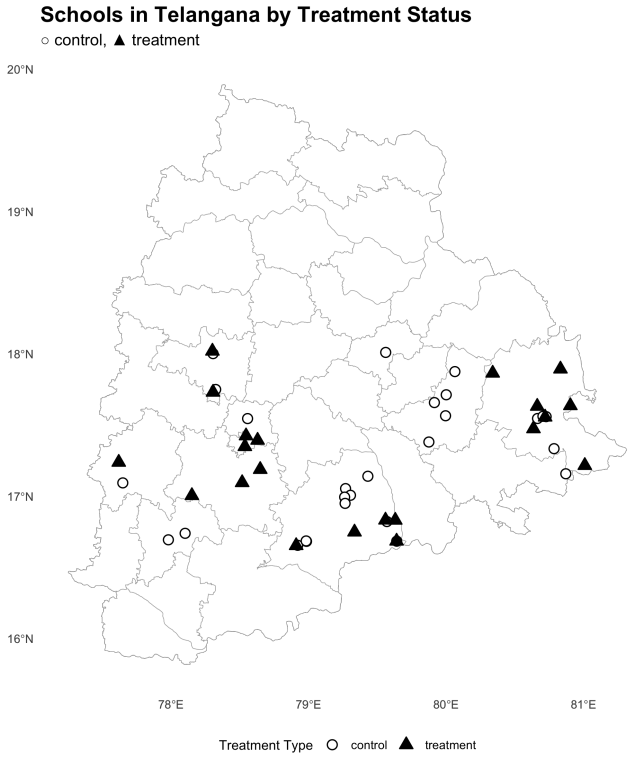


Figure 5: Geographic Distribution of Schools in Telangana by Treatment Status

Notes: The map shows the locations of schools across different districts in Telangana, with white circles representing schools receiving only the Challenge intervention and black triangles representing schools receiving both Training and Challenge interventions.

Table 7: Baseline Balance Among Endline Sample

Variable	Control Mean	Treatment Mean	Diff.	SE SE	p-val
Academic Measures					
Math Score	42.49	41.68	-0.81	2.08	0.698
Science Score	53.67	53.86	0.19	1.74	0.915
Innovation Score	45.50	46.01	0.51	1.50	0.735
Fluid Intelligence	58.75	58.93	0.17	1.71	0.919
Emotional Intelligence	44.69	44.22	-0.46	0.91	0.612
Risk Index	-0.01	0.02	0.02	0.09	0.781
Big Five					
Extraversion	25.10	24.98	-0.12	0.23	0.585
Agreeableness	25.82	25.52	-0.30	0.30	0.329
Conscientiousness	29.04	28.50	-0.54	0.36	0.138
Neuroticism	22.96	22.81	-0.15	0.18	0.421
Openness	31.92	31.70	-0.23	0.48	0.641
Socioeconomic Indicators					
SES Index	2.26	2.39	0.14	0.07	0.065
Computer Skills	0.36	0.40	0.04	0.03	0.258
Differential Selection Tests					
Joint F-test for baseline balance: $F(13, 79) = 1.06$, $p = 0.408$					
Test for differential selection: $\chi^2(3) = 0.25$, $p = 0.969$					

Notes: This table presents tests for balance in baseline characteristics among students who were present at endline data collection. The Control Mean and Treatment Mean columns show the average values for control and treatment group students respectively. The Difference column shows the estimated difference between treatment and control groups. Standard errors (SE) are clustered at the school level. The joint F-test examines whether all baseline characteristics jointly predict treatment status among endline attendees. The differential selection test examines whether baseline academic and innovation measures predict attrition differently across treatment and control groups.

Table 8: Robustness of Treatment Effects to Attrition Adjustment

Estimation Approach	Innovation Score	Math Score	Science Score
Endline Sample	1.235	0.052	1.258
	(0.784)	(1.776)	(1.846)
	[0.119]	[0.977]	[0.497]
IPW	1.255	-0.039	1.258
	(0.784)	(1.752)	(1.846)
	[0.114]	[0.982]	[0.497]
Balanced Panel	0.921	1.258	1.258
	(0.916)	(1.846)	(1.846)
	[0.318]	[0.497]	[0.497]
Observations			
Endline Sample	4,552	4,552	4,552
Balanced Panel	3,468	3,468	3,468

Notes: This table presents treatment effect estimates under three different approaches to address potential attrition bias. The Endline Sample column shows results from standard analysis on students present at endline. The Inverse Probability Weighting (IPW) column shows results using weights based on predicted endline attendance. The Balanced Panel column restricts analysis to students present in all three data collection waves. Standard errors clustered at the school level are in parentheses, and p-values are in square brackets.

Appendix

A. Qualitative Analysis of Student Feedback

To complement the quantitative findings, I collected open-ended feedback from students who participated in the innovation training program. This qualitative evidence provides additional insights into students' subjective experiences and perceived benefits of the intervention.

A.1 Data Collection and Analysis Approach

At endline, I asked treated students to provide written feedback about their experience with the Inqui-Lab program. This resulted in 2,524 text entries, of which 1,746 contained substantial responses (text length $>$ 50 characters). To analyze this large corpus of qualitative data systematically, I employed both computational sentiment analysis and thematic coding approaches:

1. **Sentiment Analysis:** Using GPT-4's natural language processing capabilities, I classified each substantive response as positive, neutral, or negative. This analysis revealed overwhelmingly positive sentiment, with 97% of responses coded as positive, 3% as neutral, and 0% as negative.

2. **Thematic Analysis:** Using computational assistance from GPT-4, I identified recurring themes in student responses. Five dominant themes emerged: problem-solving confidence, community responsibility, positive change, sense of possibility, and career aspirations.

A.2 Key Themes from Student Feedback

Representative quotes from each theme provide deeper insight into students' experiences:

Problem-Solving Confidence: *"I liked the Think and Make program from the Inqui-Lab Foundation because this is highly beneficial for the country's development. Additionally, children can learn to take more initiative in problem-solving. As a student leader, I gained valuable experience in teaching problem-solving methods. This program should achieve even greater success."*

This theme aligns with the quantitative findings on improved problem-solving abilities (0.19 SD). Students frequently mentioned enhanced confidence in approaching complex problems and applying structured thinking processes.

Community Responsibility: *"I have learned a lot from the Think and Make program. This is so interesting and exciting. In this Think and Make program, we have created 2 prototypes. I have also learned how to solve the surrounding problems and how to overcome them..."*

Students consistently referenced the application of innovation skills to address community challenges, reflecting the pattern observed in innovation content analysis where students directed attention to locally relevant problems.

Positive Change: *"I am very happy with this Inqui Lab program in 8th class. I feel delighted as it has improved my thinking abilities significantly. Many activities are taking place, and I am excited about the positive changes occurring in our school. We are advocating against plastic usage and encouraging the use of steel bottles. Similarly, we are promoting water conservation, preventing food wastage, and emphasizing the importance of saving plants. I am extremely happy about all of the changes happening around me."*

This theme emphasizes the tangible changes students implemented in their immediate environments, providing evidence that innovation training translates into real-world actions beyond classroom activities.

Sense of Possibility: *"I liked this program because I think this is very innovative, and am interested in this program. I solved so many problems through this Think and Make program. This is really useful for our life, and I learn if we have any problem in our areas or surroundings, we can easily solve the problem, so I thank to Think and Make program."*

Many students expressed a newfound sense of agency and possibility, suggesting that the program may foster attributes related to self-efficacy and growth mindset.

Career Aspiration: *"I will find some problems in our house, gurukulam, school in my surroundings, and then find their solutions. Inqui-lab's Think and Make program make me so happy. I will become a scientist in my life thank-you for this giving opportunity sirs and madams."*

This theme connects to the quantitative findings on career aspirations, particularly the increase in teacher career aspirations observed in the main results. Students articulated specific future career goals related to innovation, problem-solving, and scientific pursuits.

The consistent patterns across hundreds of independent responses provide complementary evidence for the mechanisms through which the innovation training program influenced student outcomes. The emphasis on problem-solving, community relevance, and expanded sense of possibility aligns with the theoretical framework of skill development and innovation focus discussed in the main text.

B. Additional Outcome Measure Details

B.1 Multi-Armed Bandit Task Implementation

The multi-armed bandit task was implemented as a business simulation with three distinct ventures: Fresh Juice Stand, Ice Cream Stand, and Pani Puri Stand. This implementation draws on March (1991b)'s framework that characterizes innovation as requiring a balance between exploitation (refining existing ideas) and exploration (discovering new possibilities).

Following Ederer and Manso (2013a), the task was designed so that different business choices represented different risk-reward profiles. The Fresh Juice Stand represented an exploitation strategy (lower risk, lower potential reward), while the Ice Cream Stand

and Pani Puri Stand represented exploration strategies with varying degrees of risk and potential reward.

Table 9: Probability Distribution of Profits by Business Type

Profit ( )	Probability (%)		
	Fresh Juice Stand	Ice Cream Stand	Pani Puri Stand
0	25.73	63.33	54.00
10	9.77	-	9.60
20	10.90	-	11.50
30	11.40	-	-
40	11.10	8.53	10.60
50	9.97	-	-
60	8.17	-	8.93
70	6.00	-	7.00
80	3.83	-	6.77
90	2.03	-	-
100	0.83	5.20	4.47
110	0.23	-	-
120	0.03	2.43	-
130	-	-	3.40
140	-	1.00	-
160	-	1.87	0.27
180	-	-	0.03
190	-	0.80	-
220	-	0.23	-
250	-	0.03	-
Expected Value	43.27	32.73	37.77
Std. Deviation	25.65	52.40	40.83

As shown in Table **??**, the Fresh Juice Stand (exploitation arm) had a 25.7% probability of zero profit and a maximum potential profit of 110 per hour, with relatively low variance in outcomes. The Ice Cream Stand (high-risk exploration arm) had a 63.3% probability of zero profit but a maximum potential profit of 250 per hour, with high outcome variance. The Pani Puri Stand (moderate-risk exploration arm) had a 54.0% probability of zero profit and a maximum potential profit of 180 per hour, with intermediate outcome variance.

For each business, students made decisions on four input variables: location, menu items, working hours, and price. These variables combined to influence the probability distributions, though the exact mapping was not disclosed to participants. The available input choices were:

- **Stand Type**: Fresh Juice, Pani Puri, or Ice Cream
- **Location**: In front of Market, Homes, or School

- **Working Hours**: 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 hours
- **Menu Items**: 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 items
- **Selling Price**: 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20

Students participated in 12 rounds of decision-making, with each round representing a business day. After the first round, previous choices were pre-selected as the default for subsequent rounds to provide a reference point, but students were free to change any or all inputs. An example sequence of decisions and outcomes is shown in Table **??**.

Table 10: Example Decision Sequence in Multi-Armed Bandit Task

Round	Type of Stand	Location	Working Hours	Menu Items	Price	Profit
1	Pani Puri	School	8hrs	12	13	20
2	Ice Cream	Market	12hrs	7	16	70
3	Fresh Juice	Homes	9hrs	6	13	40
4	Ice Cream	School	10hrs	11	16	0
5	Ice Cream	School	11hrs	5	13	40
6-12	Subsequent rounds...					

After each round, students received immediate feedback on their profit, which was drawn from the probability distribution associated with their chosen business type and modified by their input selections. The interface displayed their current business selection, input choices, resulting profit, and a history of previous decisions and outcomes to facilitate learning across rounds.

I measured three key aspects of innovative behavior using this task:

1. **Initial Exploration**: Whether a student began with an exploration strategy (Ice Cream or Pani Puri Stand) rather than the exploitation strategy (Fresh Juice Stand)
2. **Exploration-Exploitation Balance**: The number of times students chose to explore versus exploit across all rounds
3. **Innovation Success**: The cumulative profit achieved across all rounds, indicating ability to effectively balance exploration and exploitation to discover high-value solutions

The task was programmed using JavaScript and implemented via the offline Qualtrics app on tablets. Prior to implementation in the study, the task was piloted with 35 students from similar backgrounds to ensure appropriate difficulty levels and engagement. Pilot testing confirmed that students understood the task and demonstrated meaningful variation in strategies, with approximately 60% initially choosing exploration and 40% choosing exploitation.

This implementation allowed me to assess students' propensity for exploration versus exploitation in a controlled environment with clear payoff structures, providing a behavioral measure of innovation-relevant decision-making that complements the other innovation measures in the study.

B.2 Neurological Basis of Exploration and Exploitation

The exploration-exploitation framework has a well-established neurological basis. Daw et al. (2006) used functional magnetic resonance imaging (fMRI) to investigate neural activity during a multi-armed bandit task and found that exploratory choices were associated with increased activity in the frontopolar cortex and intraparietal sulcus, regions associated with cognitive control and decision-making under uncertainty. In contrast, exploitative choices were associated with activity in the ventromedial prefrontal cortex and ventral striatum, regions associated with reward processing and value-based decision-making. This neurological distinction supports the idea that exploration and exploitation represent fundamentally different cognitive processes.

Additional research by Laureiro-Martínez et al. (2015) examined the neural correlates of exploration and exploitation in experienced business professionals. They found that exploration was associated with increased activity in the right frontopolar cortex, while exploitation engaged the bilateral ventral striatum. Importantly, individual differences in neural activity predicted behavioral tendencies toward exploration or exploitation. These neurological findings support the theoretical framework developed by March (1991b) and provide a biological basis for measuring innovative ability through tasks that require balancing exploration and exploitation. By engaging these distinct neural systems, multi-armed bandit tasks can tap into the cognitive processes that underlie innovative thinking.

During adolescence, the brain regions associated with exploration and exploitation undergo significant development. The prefrontal cortex, critical for exploration, continues to develop into early adulthood, suggesting that interventions during this period might be particularly effective in shaping exploratory tendencies. This neurodevelopmental perspective provides additional rationale for targeting innovation training during adolescence.

C. Additional Outcome Measure Details

C.1 Innovation-S Scale Development Process

The development of the Innovation-S scale followed a rigorous multi-stage process. After conducting in-depth interviews with 14 successful innovators, I identified key themes and potential scale items. These were then refined through cognitive interviews to ensure clarity and relevance. The final survey was administered to 44 experienced innovators representing diverse professional backgrounds. The industry distribution included education (40.9%), technology (15.9%), healthcare (9.1%), finance (6.8%), manufacturing (4.5%), and others (22.8%). Participants held various roles including founders/CEOs (25.0%), product managers (13.6%), analysts/researchers (11.4%), educators (20.5%), and engineers (9.1%).

The initial pool contained 15 potential scale items, which factor analysis reduced to a 7-item scale with a two-factor structure. Factor 1, labeled "Novelty," included items such as "Offers a new solution to a problem," "Applies existing knowledge in novel ways," "Challenges conventional approaches," and "Introduces a new perspective or approach." Factor 2, labeled "Value-Add," included items such as "Improves the lives of those who will use it," "Addresses a clear need or pain point," and "Produces meaningful, measurable benefits."

The scale demonstrated acceptable internal consistency (Cronbach's alpha = 0.69) and good construct validity. Principal component analysis yielded a Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy of 0.57 and Bartlett's test of sphericity was significant ($p < 0.001$), supporting the factor structure. Validation against real-world outcomes showed strong correlations with reported user base ($r = 0.71$, $p < 0.001$) and funding raised ($r = 0.31$, $p < 0.05$), indicating good predictive validity. These associations held across innovations of varying scales, from those serving a few dozen users to those reaching millions, and funding levels ranging from a few thousand to several million dollars.

C.2 Fundraising Challenge Implementation Details

The inter-school fundraising challenge was designed to simulate real-world funding processes for innovations, providing an externally valid measure of innovation quality beyond laboratory assessments. From the 1,200 innovation ideas collected across treatment (T) and control (C) groups, I selected the top 40 ideas from each condition based on their Innovation-S scale ratings by independent evaluators. This selection process ensured that only the highest quality ideas from each group competed in the challenge, while maintaining balanced representation between treatment and control conditions.

To minimize preparation bias and simulate realistic innovation pitching conditions, teams received notification only 5 days before the event. This short notice period prevented differential preparation advantages and better reflected real-world innovation competitions where teams must quickly prepare presentations from existing work. Team size was standardized to 3-4 students per team, with members selected from the original innovation teams that developed each idea. All teams received identical presentation templates and guidelines: 5-minute pitch time, standardized slide format (problem statement, solution description, prototype details, potential impact), and 3-minute Q&A session with judges. Teams presented in randomized order, with treatment and control teams interspersed to prevent order effects. Judges were blind to treatment condition. Judges used a standardized scoring rubric using the innovation scale and funding they want to grant. Each dimension was rated on a 1-5 scale. Each judging pair had a fixed budget of 20,000 to allocate across 20 teams, with no minimum or maximum per team, requiring judges to make relative value judgments similar to real investment decisions.

The expert jury comprised eight professionals with extensive experience evaluating early-stage innovations. This included three non-profit entrepreneurs who were founders or senior leaders at established social enterprises with 5-10 years of experience implementing innovation-based solutions in education, healthcare, and rural development. Two social impact investors participated as investment managers from social impact funds with experience evaluating early-stage ventures in the 5-50 lakh (approximately $6,000-$60,000) funding range. Additionally, three corporate innovation leaders served as jurors, representing directors or managers of innovation departments at major Indian companies, responsible for evaluating internal and external innovation proposals. Jurors were recruited through professional networks and selected based on criteria including: (1) minimum 5 years of experience evaluating innovations, (2) familiarity with adolescent innovation competitions or education initiatives, and (3) no prior relationship with the research project or implementing partners to ensure impartiality.

I implemented a double-blind assessment protocol to minimize potential biases. Judges were not informed about the research design or which teams belonged to treatment vs.

control conditions. Each pair of jurors evaluated exactly 20 teams, with teams balanced across treatment and control groups (10 from each condition). Judges were randomly assigned to team groupings, and the assignment of judges to teams was stratified to ensure balanced expertise across innovation domains. Judges independently scored each presentation before conferring on funding allocations, providing both individual assessments and collective funding decisions.

The primary outcome measure from this challenge was the treatment-control difference in average fundraising amounts, representing the real-world valuation gap between innovations from treatment and control students. Secondary measures included average scores on individual assessment dimensions and qualitative feedback from judges on innovation quality. This multi-faceted approach to measuring innovation quality through expert assessment and funding allocation provided externally valid evidence on the effectiveness of the innovation training program, complementing the more controlled laboratory and survey-based measures.

C.3 Measurement Adaptation Procedures

The original RMET, developed by Baron-Cohen et al. (2001), contains 36 black and white photographs of the eye region of Caucasian adults, with each image accompanied by four emotion words. To ensure cultural relevance and age appropriateness for Indian adolescents, I created an adapted version that retained the 12 highest-discrimination items from the original test and added 8 new items featuring South Asian faces. This adaptation process included ensuring age-appropriate vocabulary for the emotion words and validating translation of emotion terms with bilingual experts. The cultural adaptation approach follows best practices established by Vellante et al. (2013) for cross-cultural application of emotional intelligence measures. The adapted version was piloted with 148 out-of-sample students from the target population, achieving a Cronbach's alpha of 0.71, comparable to the original test's reliability in adolescent populations. Item-total correlations ranged from 0.23 to 0.58, indicating good discrimination.

I used selected items from Raven's Advanced Progressive Matrices (Set II) to create a 20-item assessment appropriate for the target population. Items were selected based on increasing difficulty to minimize ceiling effects, coverage of different pattern types (e.g., addition/subtraction, figure progression), and results from pilot testing with 50 students from similar schools. The adapted version achieved a Cronbach's alpha of 0.82 in pilot testing, with good discrimination across the ability spectrum of our target population.

The science assessment was developed through a systematic process that included review of state curriculum and textbooks for grades 8-10, identification of key concepts across physics, chemistry, and biology, development of 30 multiple-choice and 5 short-answer questions, expert review by 3 science teachers from non-participating schools, and pilot testing with 65 students from similar backgrounds. The final assessment included 25 items covering fundamental scientific principles, with a Cronbach's alpha of 0.76 and appropriate difficulty distribution (mean score 58% in pilot sample).

D. Midline Results and Progression Over Time

I collected midline data after the first year of the intervention to examine how treatment effects evolved over the course of the program. Comparing midline and endline results

reveals an interesting progression that provides insight into skill development trajectories.

At midline, the innovation program had already produced significant improvements in students' innovation scores (2.03 percentage points, $p<0.05$) and related behaviors, including exploitative approach (4.82 percentage points, $p<0.01$) and explorative approach (2.55 percentage points, $p<0.05$). However, midline results showed a negative effect on "Started With Exploration" (-0.031, $p<0.05$), which reversed to a positive effect by endline (0.060, $p<0.01$). This suggests that students initially approached the innovation tasks conservatively but developed more sophisticated innovation strategies by the end of the two-year program.

A particularly striking contrast appears in the higher-order skills domain. At midline, the program showed no significant positive effects on fluid intelligence (-1.75 percentage points, not significant) and a negative effect on emotional intelligence (-1.72 percentage points, $p<0.05$). By endline, these patterns had reversed, with significant positive effects on fluid intelligence (2.63 percentage points, $p<0.05$) and problem-solving (3.67 percentage points, $p<0.05$). This progression suggests that cognitive benefits from the innovation training emerged only after sustained engagement with the program, consistent with research on neuroplasticity indicating that cognitive changes often require extended practice.

The academic outcome patterns also evolved meaningfully over time. At midline, the program had a negative effect on math scores (-2.82 percentage points, $p<0.01$) and math interest (-0.16, $p<0.01$). By endline, these negative effects had disappeared in the full sample, with math scores showing a small, non-significant positive effect (0.16 percentage points). This pattern suggests that initial trade-offs between innovation activities and traditional academics diminished over time, possibly as students learned to integrate these different forms of learning.

Student engagement metrics showed similar evolution. While attendance showed no significant treatment effect at midline (-0.17 days, not significant), by endline there was a marginally significant positive effect (1.78 percentage points, $p<0.10$). Teacher career aspirations followed a similar pattern, with no effect at midline but a significant positive effect by endline (2.1 percentage points, $p<0.01$).

These temporal patterns provide important context for interpreting the main results. The contrasting midline and endline effects suggest that some skills require sustained engagement to develop, while others may involve initial trade-offs that diminish over time. The progression also indicates that the second year of the program was crucial for developing higher-order cognitive skills and resolving potential tensions with traditional academic performance.

Table 11: Treatment Effects on Student Outcomes in Year 1

Category	Outcome	Control Mean	Treatment Effect	R^2
Panel A: Innovation Outcomes				
Innovation Measures	Innovation Score	57.66	2.029**	0.033
			(0.714) [0.011]	
	Exploitative Approach	31.80	4.820***	0.063
			(1.587) [0.006]	
	Explorative Approach	64.35	2.550**	0.029
			(1.126) [0.031]	
	Started With Exploration	0.43	-0.031**	0.013
			(0.015) [0.038]	
Panel B: Higher-Order Skills				
Cognitive and Non-cognitive Skills	Fluid Intelligence	54.31	-1.748	0.074
			(1.283) [0.233]	
	Emotional Intelligence	44.16	-1.721**	0.094
			(0.750) [0.035]	
	Big 5: Extraversion	3.21	-0.035**	0.042
			(0.014) [0.019]	
	Big 5: Agreeableness	3.32	-0.002	0.079
			(0.017) [0.931]	
	Big 5: Conscientiousness	3.23	-0.017	0.058
			(0.015) [0.278]	
	Big 5: Neuroticism	2.85	0.035**	0.011
			(0.014) [0.028]	
	Big 5: Openness	3.25	0.015	0.104
			(0.020) [0.522]	
Panel C: Academic Outcomes				
Academic Performance and Engagement	Math Score	45.30	-2.819***	0.175
			(0.917) [0.013]	
	Science Score	50.92	0.394	0.157
			(0.662) [0.541]	
	Liked Math	2.55	-0.158***	0.018
			(0.034) [0.000]	
	Liked Science	2.50	-0.044	0.008
			(0.039) [0.307]	
Panel D: Survey Responses				
Student Behavior and Aspirations	Attendance	10.36	-0.168	0.011
			(0.149) [0.315]	
	Questions Asked	3.18	-0.186*	0.014
			(0.093) [0.059]	
	Number of Friends	4.71	-0.129	0.050
			(0.162) [0.448]	
	STEM Career Aspiration	0.33	0.002	0.197
			(0.014) [0.913]	
	Teacher Career Aspiration	0.06	0.011	0.153
			(0.008) [0.158]	
Panel E: Risk Behaviors				
Risk Measures	Risk Index	0.05	-0.092	0.019
			(0.074) [0.265]	
	Super-betting	3.21	-0.109	0.011
			(0.074) [0.169]	

Notes: Each cell reports the coefficient on the treatment indicator from a separate regression.
Values in parentheses are standard errors, clustered at the school level.
The value in square brackets is the wild bootstrap p-value, calculated using 1000 replications.
Statistical significance is denoted by $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.
All regressions include strata fixed effects, the outcome's baseline value when available, demographic controls, and school characteristics.
Control mean reports the mean value for the outcome variable in the control group in Year 1 (midline).

E. Power Calculation Details

I used the formula for cluster-randomized trials as described by Dong and Maynard (2013), which incorporates key parameters such as the desired significance level, desired power, proportion of variance explained by covariates, intraclass correlation coefficient, average cluster size, proportion of clusters assigned to treatment, and total number of clusters. My calculations were based on the following assumptions: 80 schools (40 treatment, 40 control) as per the actual study design; an average of 50 students per school, which is a conservative estimate based on expected enrollment and potential attrition; an Intraclass Correlation Coefficient (ICC) of 0.15, which is a typical value for educational studies (Hedges and Hedberg, 2007); an R-squared from baseline covariates of 0.3, which is a conservative estimate based on similar studies in educational settings; a desired power of 80%, which is a commonly accepted standard in social science research; a significance level of 5% (two-tailed), which is the conventional level for hypothesis testing; and an equal allocation to treatment and control groups. Using these parameters, I calculated the Minimum Detectable Effect Size (MDES) as follows:

$$
\begin{aligned}
&\text{MDES} = (t_{\alpha/2} + t_{\beta}) * \sqrt{\frac{(1-R^2)(1+(m-1)ICC)}{P(1-P)Nm}} \\
&= (1.96 + 0.84) * \sqrt{\frac{(1-0.3)(1+(50-1)0.15)}{0.5(1-0.5)80*50}} \\
&= 2.8 * \sqrt{\frac{(0.7)*(8.35)}{2000}} \\
&= 2.8 * \sqrt{0.0029225} \\
&= 2.8 * 0.0540601 \\
&= 0.1514
\end{aligned}
$$

This calculation indicates that the study design has 80% power to detect an effect size of 0.1514 standard deviations at the 5% significance level (two-tailed). This effect size is considered educationally meaningful, as effect sizes of 0.15-0.20 are often viewed as substantively important in educational interventions (Hill et al., 2008). It's worth noting that the inclusion of baseline outcome measures as covariates in the final analysis is expected to further improve statistical power (McKenzie, 2012), potentially allowing for the detection of even smaller effects. Additionally, the stratified randomization design may contribute to increased power by reducing between-cluster variation. Potential limitations of this power calculation include assumptions about cluster size and the potential for differential attrition. Actual power may vary depending on the realized sample size and intraclass correlation.