Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1660

Can Innovation be Taught in Schools? Experimental Evidence from India

Saloni Gupta
Brown University

Childhood exposure to innovation shapes who becomes an inventor, but most children grow up without it. This study uses a randomized evaluation across 80 public schools in India to show that schools can provide part of this exposure. A curriculum teaching the innovation process raised participation in an innovation competition, the quality of student ideas, and skills in exploration, idea generation, fluid intelligence, and problem solving. State graduation exam scores explain at most 10 percent of skill variation, suggesting that STEM track selection based on academic scores alone may overlook innovation-relevant talent. The program cost about $40 per student annually.

VERSION: November 2026

Suggested citation: Gupta, Saloni. (2026). Can Innovation be Taught in Schools? Experimental Evidence from India. (EdWorkingPaper: 26-1660). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/gbxp-1327

Can Innovation be Taught in Schools?

Experimental Evidence from India

Saloni Gupta*

Abstract

Childhood exposure to innovation shapes who becomes an inventor, but most children grow up without it. This study uses a randomized evaluation across 80 public schools in India to show that schools can provide part of this exposure. A curriculum teaching the innovation process raised participation in an innovation competition, the quality of student ideas, and skills in exploration, idea generation, fluid intelligence, and problem solving. State graduation exam scores explain at most 10 percent of skill variation, suggesting that STEM track selection based on academic scores alone may overlook innovation-relevant talent. The program cost about $40 per student annually.

JEL Codes: I21, I25, O31, O15, J24

Keywords: Innovation, education, government, India, higher-order skills, scale

*Gupta: Brown University, Department of Education (email: saloni_gupta@brown.edu). AEA RCT Registry: AEARCTR-0009696 (Gupta 2025). 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, Eshwar Bandi, Alex Eble, Judith Scott-Clayton, Anjali Adukia, Andreas de Barros, Carolina Concha-Arriagada, Ashutosh Bhuradia, Sidra Ali, Sandra Black, Abhijeet Singh, Guthrie Gray-Lobe, and Michael Best for their feedback and support at multiple stages of this project. I have also benefited from presenting this work at academic venues including Stanford University, Brown University, The University of Chicago, Columbia University, Helsinki Graduate School of Economics, Paris School of Economics, Ashoka University, University of Zagreb, and the Advances in Field Experiments and Association for Education Finance and Policy conferences. This project was financially supported by the Weiss Fund at the University of Chicago and multiple small research grants from Teachers College, Columbia University. I gratefully acknowledge the NAEd/Spencer Dissertation Fellowship, which supported my work on this paper. This study received IRB approval from Teachers College, Columbia University (Protocols #22-352 and #22-310), and Stanford University (Protocol #79854). AI software (Claude, Anthropic; Codex, OpenAI) assisted during manuscript preparation. The author is solely responsible for all content.

I. Introduction

Long-run economic growth depends on the production of new ideas (Romer 1990). Participation in this process is unequal: the probability of becoming an inventor is shaped in part by childhood exposure to innovation and varies sharply by family income, gender, and geography (Bell et al. 2019; Hoisl, Kongsted, and Mariani 2023; Aghion, Akcigit, Hyytinen, and Toivanen 2017).[1] Such childhood exposure may make inventive careers visible, show how ideas are developed, and create opportunities to engage in that process. Extending this exposure to more children is particularly relevant in low- and middle-income countries, where formal research workforces are much smaller (Jones 2022). India reported 259 full-time-equivalent R&D researchers per million people in 2020, compared with 490 in lower-middle-income economies and 4,149 in high-income economies (World Bank 2024b, 2026b). Schools cannot reproduce the relationships and environments that ordinarily provide exposure to innovation, but their broad reach may allow them to offer one component at scale: repeated, guided practice in developing and testing ideas.

This paper asks whether such instruction builds students' skills in creating early-stage innovations.[2] I study this question in a randomized controlled trial across 80 government residential schools in Telangana, India, a state that combines a global technology hub in Hyderabad with a large agricultural workforce (Government of Telangana 2024). The schools serve students from Scheduled Castes and Scheduled Tribes, historically marginalized caste and tribal communities, whose parents predominantly work as farmers or daily-wage laborers in the informal sector. Of the 80 schools, 40 received the program and 40 served as active controls. The study follows approximately 5,900 grade 8 students for three academic years, through the high-stakes grade 10 state graduation examinations. The program, implemented by the Government of Telangana in partnership with the Inqui-Lab Foundation, comprised four instructional cycles over two years. In each cycle, students identified problems in their communities, interviewed the people affected, generated and compared possible solutions, and finally built and revised prototypes using contemporary technology and material kits. Two trained student leaders in each classroom led the sessions using structured workbooks, and

[1] Bell et al. (2019) isolate childhood exposure effects using the tendency of children to patent in the same narrowly defined technology classes as inventors in their families and neighborhoods. The specificity of these patterns is difficult to explain by inherited differences in ability or preferences alone. Bell et al. (2019) and Aghion, Akcigit, Hyytinen, and Toivanen (2017) also show that children from higher-income families are more likely to become inventors in the United States and Finland, respectively. Aghion, Akcigit, Hyytinen, and Toivanen (2023) identify a causal effect of parental education on children's likelihood of becoming inventors and provide suggestive evidence that Finland's comprehensive-school reform improved those chances for children of less-educated parents. Across countries, Agarwal and Gaulé (2020) show that, conditional on International Mathematical Olympiad performance, participants from lower-income countries are less likely to enter mathematics PhD programs and produce less research.

[2] The Oslo Manual, the OECD and Eurostat guidelines for measuring innovation, defines an innovation as "a new or improved product or process (or combination thereof) that differs significantly from the unit's previous products or processes and that has been made available to potential users (product) or brought into use by the unit (process)" (OECD and Eurostat 2018). Student ideas and prototypes satisfy the first part of this definition but not the implementation requirement. The study emphasizes novelty and usefulness at the idea and prototype stages; *innovation* refers here to this early production.

trained teachers were present to support the student leaders. The evaluation has an active control: students in all 80 schools could submit ideas to the Innovation Challenge, which operated under the same announcements, deadlines, and prizes, and shortlisted teams from either arm were invited to pitch their ideas to an expert jury. The design therefore varied access to guided practice while holding fixed the formal opportunity to enter the competition.

Evaluating school-based instruction in innovation raises three challenges: measurement validity, relevance for students with low academic preparation, and the crowding out of preparation for high-stakes examinations. First, innovation is difficult to measure at this stage of schooling: patents and commercialization, the conventional measures of innovation, are not observable, and there is no established standardized and scalable measure of the ideas students produce. Second, students with lower foundational academic skills may benefit less from instruction in higher-order innovation skills. Dynamic complementarity in skill formation implies that skills acquired earlier can raise the productivity of later investments (Cunha and Heckman 2007), and Pritchett and Beatty (2015) show that curricula that advance faster than students' mastery can generate little learning in developing-country settings.[3] This concern is especially relevant in India, where 54 percent of rural grade 8 students cannot solve a division problem set at the grade 3 to 4 level (ASER Centre 2025). Third, the program's 51 contact hours may displace time for core academic learning, particularly because the intervention takes place during grades 8 and 9, immediately before the grade 10 high-stakes examinations.

To measure the novelty and usefulness of the ideas students produce, I construct Innovation-S by translating the OECD and Eurostat definition of innovation into seven five-point items (OECD and Eurostat 2018). The items were also built and validated using interviews with 14 practicing innovators and a survey of 44 adults across professions who build products they regard as innovative. In the study's second year, teams from treatment and control schools submitted 1,225 ideas to the Innovation Challenge using a common format (a drawing and an audio-recorded pitch), and evaluators blind to treatment status scored them with the Innovation-S scale. At the finals, professional jurors assessed the live pitches of the top 80 student teams using the same Innovation-S scale and also provided funding at their discretion. The within-juror correlation between rubric scores and funding allocations was 0.77.

I also use a second approach to measure innovation, assessing the skills that its multistage process draws on: exploring alternatives, generating solutions, diagnosing design problems, and building and testing prototypes. This complements the idea-quality measure by examining whether the program developed the underlying skills involved in producing those ideas. To measure these skills, I adapted standardized

[3] In the education literature, a common interpretation of Bloom's taxonomy treats higher-order objectives as building on lower-order ones, which raises the concern that students without foundational skills may not benefit from programs focusing on higher-order skills such as application and synthesis (Bloom et al. 1956; Anderson and Krathwohl 2001).

assessments from the research literature: for exploration skills, an incentivized search task (March 1991; Ederer and Manso 2013); for idea generation, creative-thinking tasks (OECD 2024); for fluid intelligence, the ICAR test (Condon and Revelle 2014); and for well-defined problem solving (problems with a specified goal and fixed rules for reaching it), Sudoku puzzles (Newell and Simon 1972). The last two serve as proxies for the diagnostic and problem-solving skills that prototyping draws on. Gains on these assessments, which use tasks distinct from those practiced in the curriculum, provide evidence on whether the program developed the underlying skills.

I report six main results. First, descriptively, the academic scores used to screen students into STEM education are weak predictors of innovation-relevant skills. Academic grades and examination scores are widely used to determine access to STEM education, from upper-secondary science streams to university programs. Among control students, the high-stakes grade 10 state graduation score explains at most 10 percent of the variation in idea generation, problem solving, fluid intelligence, or exploration (Figure 1). A screen limited to the top baseline or grade 10 academic quintile would miss 60 to 80 percent of the students who rank in the top quintile on these skills.

Second, more importantly, structured instruction changed both who produced innovative work and the quality of what they produced. In control schools, 73.3 percent of students participated in the end-of-year Innovation Challenge held in their school, and treatment raised participation by 8.0 percentage points. Treatment increased novelty scores by 0.187 standard deviations and usefulness scores by 0.146 standard deviations. In the final round of the Innovation Challenge, treatment schools accounted for eight of the ten teams that received the most funding from jurors. Together, these effects on participation and idea quality show that schools can teach the process of developing ideas.

Third, beyond idea quality, treatment improved four higher-order skills that are components of the innovation process. The exploration index rose by 0.094 index points, idea generation by 0.314 standard deviations, fluid intelligence by 0.111, and problem solving by 0.142. The gains show transfer beyond the curriculum's projects. Students practiced exploring alternatives, generating solutions, and building prototypes for community problems, and improved on standardized assessments of the underlying skills.

Fourth, turning to academic trade-offs, treatment produced no detectable change in academic performance. At endline, the effects on mathematics and science are -0.007 and 0.027 standard deviations and statistically indistinguishable from zero. On the grade 10 state graduation examinations a year later, for which administrative records cover nearly the entire cohort, the estimated effects on mathematics and science are statistically indistinguishable from zero, and those on the three language subjects (English, Telugu, Hindi) are positive but do not survive adjustment for multiple testing. Reported time use is consistent with these

estimates: treatment students reported no detectable reduction in the percentages of weekly time devoted to school periods and homework. Treatment students reported allocating greater percentages of weekly time to the Innovation Challenge and sports and a smaller percentage to other activities, including dance, singing, drawing, and reading.

Fifth, with regard to heterogeneity by baseline academic performance, students with lower baseline academic performance made substantial gains in participation and innovation-relevant skills. Participation rose by 9.7 to 12.0 percentage points in the bottom three academic quintiles. Idea generation improved in every quintile, with the largest gain, 0.63 standard deviations, in the top quintile. Fluid-intelligence gains were detected only in the top quintile, while exploration estimates were positive across all quintiles (Figure 5).

Sixth, with regard to the direction of ideas, girls' teams direct a larger share of their submissions toward people and care (public welfare, health, and households), while boys' teams direct a larger share toward the environment and transport.[4] Submissions from tribal-welfare and social-welfare schools also differ in sector: teams from tribal-welfare schools direct a larger share toward agriculture and transport, and teams from social-welfare schools toward the environment and household life. Submissions from treatment schools focus more on school-related problems than those from control schools, consistent with the curriculum's instruction to observe their surroundings—in these students' case, their residential schools.

This study contributes to three main bodies of literature. The first examines the skills beyond academic achievement that schools can develop and that increasingly matter for long-term outcomes. Foundational literacy and numeracy receive strong emphasis in education policy and spending in low- and middle-income countries (Evans and Hares 2021; World Bank 2024a; CGD 2025); India's National Education Policy names them the highest priority of the education system (Government of India 2020, para. 2.2). Yet economically valuable human capital extends beyond these foundations: non-cognitive, social, and higher-order skills also shape educational attainment, employment, and earnings (Heckman, Stixrud, and Urzua 2006; Heckman and Kautz 2012; Deming 2017, 2022). Schools can develop these broader skills, as interventions fostering grit (Alan, Boneva, and Ertac 2019), curiosity (Alan and Mumcu 2024), and cognitive endurance (Brown, Kaur, Kingdon, and Schofield 2025) have been shown to improve academic performance. The benefits can persist long after instruction ends, with negotiation training raising educational attainment ten years later (Ashraf, Bau, Low, and McGinn 2020; Ashraf, Bau, Low, and Shan 2025) and entrepreneurship training improving business performance nine years later (Chioda, Gertler, Contreras-Loya, and Carney 2026). Pedagogical practices that give students an active role in learning can also improve educational outcomes. Engaging

[4]The focus of girls' teams on people-focused problems is not specific to this sample. Among 327,599 submissions to the 2025 round of the national School Innovation Marathon, 58.7 percent of teams submitting in women-and-child-development themes are girl-majority, against 50.0 percent of teams overall (Inqui-Lab Foundation and Gupta forthcoming).

students in questioning, investigation, and problem solving through inquiry- and problem-based approaches improves learning (Bando, Näslund-Hadley, and Gertler 2019; Nourani, Ashraf, and Banerjee 2025), while assigning them teaching responsibilities improves their behavior and educational progression (Alan and Kubilay 2025).

Building on this work, this paper provides experimental evidence that schools can also build human capital for innovation, a central engine of economic growth (Romer 1990; Aghion and Howitt 1992). Conventional academic tasks are closed-ended, with the problem specified and a correct answer defined, whereas innovation is open-ended, with the problem itself to be formulated and no solution known in advance. This paper shows that a standardized school curriculum can also support open-ended production, in which students must define the problems themselves and develop and revise their own solutions without a predetermined answer. It further shows that government schools in low-income settings can teach a structured process for innovation without specialist teachers or dedicated laboratories, with students learning mainly from one another through group work. These findings provide an affirmative answer to the question in the title: schools can teach students to develop more novel and useful ideas.

The second contribution adds to the literature on talent for innovation. Family background and exposure to innovation shape who becomes an inventor (Bell et al. 2019; Aghion, Akcigit, Hyytinen, and Toivanen 2017; Toivanen and Väänänen 2016), and the resulting misallocation of talent is costly. Hsieh, Hurst, Jones, and Klenow (2019) estimate that improved allocation of talent accounts for 20–40 percent of growth in U.S. market output per person between 1960 and 2010. One response is to identify talent early. Teenage competition results carry a strong signal of later scientific achievement: medalists at the International Science Olympiads, about 1,000 students a year, account for 10 percent of winners of major science prizes in their sample (Agarwal, Gaulé, and Jiang 2025). While these findings highlight the value of identifying talent in the right tail of academic achievement, this study shows that innovation-relevant strengths are more broadly distributed and can also be developed: they extend beyond the top academic quintile and respond to instruction even among students with lower academic performance.

The third contribution adds to the innovation and education policy literature by providing evidence on delivery and measurement at an intermediate scale of implementation (Al-Ubaydli, List, and Suskind 2017; List 2024; Muralidharan and Singh 2025).[5] Governments spend billions of dollars each year on research and development and on startup ecosystems (World Bank 2026a; World Intellectual Property Organization 2024). The effect of this spending on innovation depends on the supply of people with the human capital

[5]Muralidharan and Singh (2025) emphasize the importance of adapting delivery to constraints likely to bind at scale and evaluating the adapted model before wider expansion. This study provides experimental evidence that one adaptation to constraints on teacher time—structured instruction led by trained students with teacher support—is effective with the training and monitoring provided.

to develop new ideas (Bloom, Van Reenen, and Williams 2019; Van Reenen 2021). Governments in India, Rwanda, Kenya, and at least six U.S. states have supported school innovation programs at substantial scale, with reported participation reaching millions of students in India and projected reach extending to tens of thousands in U.S. school programs.[6] These initiatives fund different combinations of laboratories and equipment, teacher training, student innovation projects, entrepreneurship curricula, and school-business partnerships. By showing that innovation-relevant skills can be developed even in resource-constrained government schools, this study strengthens the case for public investment in innovation education and provides evidence at an intermediate scale to guide the instructional design of these expanding initiatives.

The remainder of the paper proceeds as follows. Sections II–V describe the setting and intervention, experimental design, outcomes, and empirical strategy. Section VI presents the results, Section VII discusses cost-effectiveness and scalability, and Section VIII concludes.

II. Context and Intervention

A. Setting

The study was conducted in Telangana's government residential schools for students from Scheduled Castes (SCs) and Scheduled Tribes (STs), two constitutionally recognized groups that have experienced persistent historical disadvantage. Scheduled Castes (also known as Dalits) were historically subjected to untouchability and exclusion under the caste system, while Scheduled Tribes (also known as Adivasis) are indigenous communities that have long faced geographic isolation and social and economic marginalization. The schools belong to two parallel state networks, Social Welfare Residential Schools, which predominantly

[6]The scale of these initiatives encompasses different forms of support. In India, Atal Tinkering Labs provide school innovation laboratories; the program reports 10,000 laboratories engaging more than 11 million students (Atal Innovation Mission 2024, 2026). The School Innovation Marathon invites students to develop solutions to problems in their surroundings (Press Information Bureau 2024), while Delhi and Andhra Pradesh have introduced entrepreneurship-mindset curricula (SCERT Delhi 2021; Amaral et al. 2025). Elsewhere, Rwanda requires secondary students to take an entrepreneurship course (Blimpo and Pugatch 2021), while Kenya's grade 9 pre-technical studies curriculum includes a project addressing a local problem (Kenya Institute of Curriculum Development 2024).

U.S. state programs likewise support different combinations of infrastructure and instruction. Wisconsin's fabrication-laboratory grants financed school equipment through 242 grants totaling approximately $5.8 million to 136 districts, with the final round in FY2026 (Wisconsin Economic Development Corporation 2026). Massachusetts's Life Sciences Center combines equipment and materials with teacher professional development and, by October 2025, had awarded more than $31.8 million to over 400 public schools. Its October 2025 school awards alone covered 131 schools and were expected to reach nearly 56,000 students (Massachusetts Life Sciences Center 2025). Connecticut's Manufacturing Innovation Fund supported technical-high-school innovation education through a $54,000 contract covering staff development, student resources, and access to the competition finals (Connecticut Department of Economic and Community Development 2026). Iowa's STEM BEST grants support teachers and employers in jointly developing curricula and projects; since 2014, 262 awards have supported the creation or expansion of more than 150 models across the state (Iowa Department of Education 2026). Arkansas supports school-based community problem-solving through the EAST Initiative (EAST Initiative 2024). Kentucky funds a three-week residential entrepreneurship program for high-school students (Commonwealth of Kentucky, Office of State Budget Director 2024). These differences make the content and delivery of instruction central to interpreting what public investment in school innovation achieves.

serve Scheduled Caste students, and Tribal Welfare Residential Schools, which predominantly serve Scheduled Tribe students; the two societies that administer them operate more than 430 single-gender institutions in total, from primary schools through undergraduate colleges (TGSWREIS 2026; TGTWREIS 2026).

The households these schools serve depend largely on manual and informal work. Caste continues to shape occupational sorting, social networks, and access to credit, limiting economic opportunities across generations (Munshi 2019); caste-based wealth gaps are large and persistent (Bharti 2018), and intergenerational mobility is low, despite the gains from affirmative action (Asher, Novosad, and Rafkin 2024). Exposure to technical and inventive work, which children may otherwise acquire through family occupations or social networks, is less likely in these settings.

Within these disadvantaged communities, however, the residential schools select students on academic performance. Admission is determined by a statewide entrance examination: in the 2026 cycle, about 173,000 students took the examination and about 48,000 were selected for admission across the state's residential school societies (Government of Telangana 2026; Sihag 2026). Once enrolled, students in both networks perform well on high-stakes state assessments. In the 2022 grade 10 board examinations, pass rates were 98.1 percent in the social-welfare schools and 96.0 percent in the tribal-welfare schools, against a statewide rate of about 90 percent (TSWREIS 2022; TGTWREIS 2024); these figures reflect the selectivity of admissions as much as the schools' own contribution. By conventional academic measures, these students therefore stand above the state average.

B. The Innovation Curriculum

Treatment schools received Think & Make, a two-year structured curriculum that teaches the process of innovation. The Inqui-Lab Foundation developed the curriculum and has delivered it in these welfare schools since 2017, in partnership with the Government of Telangana. The foundation now operates in four states. The curriculum comprises four blocks over the two study years (grades 8 and 9), two per year: Planet and People in Year 1, and Design Evolutions and Smarter Everyday in Year 2. Each block is a sequence of seven to nine weekly 90-minute sessions, 51 contact hours in total. Students work through the sessions in teams of four; teams of three or five were rare. Appendix Table B.1 summarizes the four-block sequence, required student work, and classroom delivery model.

Each block follows the same five-stage cycle: find, learn, ideate, make and test, and present. Teams observe their school, homes, and public spaces and interview the people affected to select a problem (find); study how existing products and solutions evolved (learn); generate multiple solutions under time and material constraints and shortlist them against criteria (ideate); build a prototype from kit materials, gather feedback

from at least two users, and revise (make and test); and present how they developed and tested their idea to their class and school (present). Teams build and rebuild prototypes across sessions, debugging problems and refining their designs.

Each stage of the cycle both advances the team toward a working prototype and provides practice in a distinct innovation-relevant skill. Finding problems through observation and interviews practices perspective-taking; studying how designs evolve practices pattern recognition; generating and shortlisting solutions practices divergent thinking and evaluative judgment; building and revising prototypes practices structured problem solving and persistence; presenting and absorbing feedback practices communication. Repeating these demands over four cycles is the mechanism by which the curriculum is designed to build innovation-relevant skills. Inqui-Lab's session guides map each stage to the skills it is meant to build; the research team then paired each skill with an established assessment from the relevant discipline. Section IV describes this mapping further.

The curriculum becomes progressively more open-ended as students repeat the cycle. In Year 1, teams use guided story cards with step-by-step prompts and structured evaluation rubrics; in Year 2, they progress to semi-open timed design challenges and then to fully independent problem discovery, in which teams define both the problem and the solution themselves. Smarter Everyday applies the same problem-solving cycle to a new medium, sensor-based prototyping, with students wiring sound, light, heat, and distance sensors to outputs such as LEDs, motors, and buzzers. Each classroom received an innovation kit, valued at ₹15,000 (about $180), containing circuits, sensors, recyclable materials, and beginner-friendly prototyping tools. Appendix Figures B.1, B.2, and B.3 reproduce pages from the student workbooks.

The program used a student-led active-learning model. Sessions used structured student workbooks containing step-by-step prompts, reflection sheets, and peer-feedback rubrics that allowed teams to work semi-autonomously. Two student leaders per classroom, nominated by teachers, attended 20 hours of training together with teachers at the beginning of the school year. They facilitated the sessions, managed time and materials, and were instructed to answer only clarifying questions.

The choice of student-run delivery reflected the constraints of the government school system. Teachers in these schools face high student-teacher ratios and substantial administrative duties. The program therefore assigned them a limited supporting role. A teacher was expected to be present during sessions but did not deliver the curriculum. The teacher supported the student leaders as needed, mediated when problems arose, advised teams, and helped organize the end-of-block showcases. Neither the student leaders nor the teacher supplied problems or solutions; defining the problem and producing the solution remained the teams' own work. Sessions were held during the after-school study period or in place of other non-academic periods.

The curriculum did not directly teach or test mathematics, science, or other examination-subject content. Its designers consulted India's national science textbooks to keep activities within the range of material students had already encountered.

The program paired student-run delivery with regular external monitoring. Seven field-team observers covered the 40 treatment schools over the two years. They observed 627 sessions, or 46 percent of the scheduled curriculum, and used a standardized observation rubric to record classroom implementation.

Classroom observations show student leaders managing sessions and most teams completing the planned work. An observed section enrolled approximately 40 students, who formed nine to ten teams of four; observers recorded most or all teams completing the session's activities in 83 percent of observed sessions. In the Planet unit, for example, the problems teams selected and carried through the cycle were drawn from the students' surroundings, most often the waste of water and food, and local pollution. The two student leaders led the session: they opened it by reviewing the previous session's work, announced the day's goals, managed the time, circulated among the teams to encourage participation and answer clarifying questions, and closed the session.

Teachers were present in approximately two-thirds of the classroom observations for which their presence was recorded; for context, nationally representative evidence from rural Indian public schools documents a teacher absence rate of 23.6 percent during unannounced visits (Muralidharan, Das, Holla, and Mohpal 2017). Table 1 reports completion and observation by curriculum unit.

Schools completed 98.9 percent of scheduled sessions across the four blocks over the two program years. The schedule comprised 18 sessions per treatment school in Year 1 and 16 sessions per treatment school in Year 2. Across the 40 treatment schools, 1,345 of the 1,360 scheduled sessions were delivered. In each program year, 35 schools completed every session scheduled across both structured units. Completion status was compiled from progress reported by the student leaders and audits of student manuals during school visits.

The program cost approximately $40 per student per year; Section VII examines cost-effectiveness and scalability.

III. Experimental Design and Data

A. School Selection and Randomization

The study draws 80 schools from the 320 secondary schools serving grade 8 that the two school-management societies operate. The 80 schools span 14 districts and every institutional type in the system:

TABLE 1—IMPLEMENTATION AND MONITORING OF THINK & MAKE

	Year 1		Year 2		Tracked total
	Planet	People	Design Evolutions	Smarter Everyday	
Sessions scheduled per school	9	9	9	7	34
Scheduled school-session records (sessions per school × 40 schools)	360	360	360	280	1,360
Completed school-session records	359	354	357	275	1,345
Session completion rate	99.7%	98.3%	99.2%	98.2%	98.9%
Field-team observer present, *N* (% scheduled sessions)	215 (59.7%)	183 (50.8%)	146 (40.6%)	83 (29.6%)	627 (46.1%)
Schools completing every scheduled session	39/40	35/40	37/40	38/40	35/40 in each year

Notes: The table reports implementation across the 40 treatment schools. Each structured unit scheduled seven or nine sessions per school, producing the school-session totals shown in the table. Completion status was compiled from progress reported by the student leaders and observers' reviews of student manuals during school visits. Seven field-team observers covered the 40 schools over the two program years. The observer row reports sessions at which a field-team observer was present. In each program year, 35 schools completed every session scheduled across both structured units.

social-welfare and tribal-welfare management, girls-only and boys-only composition, and rural and urban locations.[7] Figure 2 summarizes the research design and timeline.

Before the baseline assessment, I randomized schools within eight strata defined by management, gender composition, and location. Half of the schools were assigned to treatment and half to control. School characteristics were balanced across arms (Appendix Table A.2).

B. The Innovation Challenge and the Control Condition

All 80 study schools held the Innovation Challenge, a team competition, at the end of each school year. The two management societies issued identical circulars to treatment and control schools, with the same deadlines and prizes. The announcements stated that selected teams would be invited to Shark Tank-style finals in the state capital, where they could win funding (Section IV). For these residential-school students, most in rural areas, the trip to the capital was part of the opportunity. The format was familiar: *Shark Tank India* had aired three national seasons before the study's Year 2 finals,[8] and Indian business press describes the show as encouraging young people to pursue entrepreneurship (Khan 2024).

Control schools received administrative encouragement to enter the competition but no instruction in developing ideas. Teams in control schools were announced at the beginning of the first school year. Students were asked to submit ideas to the Innovation Challenge with the same team at the end of both years. Headmasters and teachers were informed through WhatsApp that surveyors would visit at the end

[7]The study sample was selected using two criteria: schools were located in districts that Inqui-Lab had shortlisted as feasible for its field team's in-person support, and they had no other externally run program active when the study began.

[8]Shark Tank is a televised pitch format in which entrepreneurs present business ideas to a panel of investors who may fund the venture; the Indian version debuted in 2021 and has aired five national seasons (SonyLIV 2026; Jha 2025).

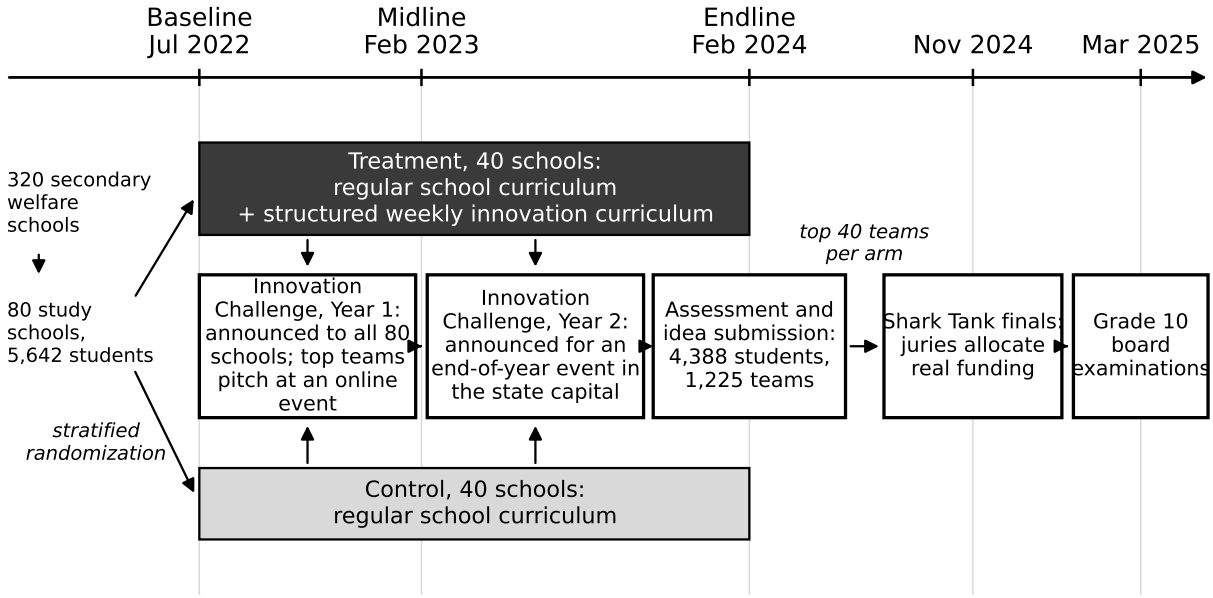


FIGURE 2. RESEARCH DESIGN AND TIMELINE

Notes: The figure shows the study timeline for the 80 study schools, drawn from 320 secondary welfare schools and randomized within strata into 40 treatment and 40 control schools. The Innovation Challenge was announced to all 80 schools at the start of each school year; top-rated teams from both arms pitched at an online event at the end of Year 1. Idea submissions are measured for the Year 2 cycle and were collected during the endline school visit, in which 4,388 students completed the individual endline assessment and 1,225 teams submitted ideas. The top 40 teams from each arm were invited to the Shark Tank finals, where juries allocated funding among the ideas they heard. Grade 10 board examinations followed in March 2025. The horizontal axis is not drawn to scale.

of each school year and that every team would be asked to submit an idea, and the circulars and WhatsApp messages announced the criteria on which ideas would be judged, so control teams were informed of the assessment criteria. Teams prepared their submissions with the support available from their teachers and headmasters. Control schools did not receive the innovation kits during the program years. Once the finalists were announced, the 40 control finalist teams received the same kits. Finalist teams in both arms had two weeks to prepare, matching the duration of the curriculum's prototype module. School innovation competitions are an established policy instrument in India: Telangana's School Innovation Challenge, run with UNICEF since 2020, and the national School Innovation Marathon, launched in 2024 and followed in 2025 by the Viksit Bharat Buildathon, invite students to present solutions to practical problems (Telangana State Innovation Cell 2026; Dejuan and Sanz Navarro 2023; Press Information Bureau 2024, 2025). The control condition therefore approximates the prevailing low-cost model for encouraging innovation in schools: an announced competition, official encouragement, and reliance on existing school staff.[9]

[9]The research team did not tell control schools about the Inqui-Lab program in the other 40 schools, but headmasters may have learned through the societies' administrative networks that Inqui-Lab was delivering a program in other schools. Such awareness could have increased or reduced effort in control schools; the design does not separately identify these responses.

C. Data Collection

I administered individual tablet assessments at baseline, midline, and endline. The baseline, administered before the program began, measured academic achievement, fluid intelligence, emotional intelligence, risk preferences, Big Five personality, and exploration behavior on the incentivized search task. The midline, administered in February 2023 near the end of Year 1, repeated the baseline instruments. The endline, at the end of Year 2, administered new variations of the same instruments and added creativity and problem solving, together with the Innovation Challenge. The Challenge was conducted in both program years, but only the Year 2 submissions are scored. The Year 1 submissions proved unusable at similar rates in both arms,[10] and the protocol was redesigned for Year 2 to capture the students' own audio-recorded pitches. Appendix Table A.1 reports the number of students assessed in each wave.

At endline, 1,225 teams submitted ideas to the Innovation Challenge on the day of the individual assessment. The individual assessment yields the student-level outcomes; the team submissions yield the participation and idea-quality outcomes, defined in Section IV. Appendix Table A.1, Panel B, reports the division of the 4,388 endline-assessed students into submitters and non-submitters and the construction of the team analysis sample. Students were assigned to teams of four at random after baseline. Most teams had reorganized by the Year 2 competition, and the share of assigned teams with unchanged membership was similar across arms (22.9 percent treatment, 21.9 percent control, $p = 0.71$). Team composition at endline is therefore largely self-selected in both arms.

Across the individual assessments and team submissions, data collection extended beyond multiple-choice questions to include open-ended text responses, interactive tasks, and audio-recorded pitches in Telugu, the students' mother tongue.[11] In this study, tablets collected students' written responses and delivered the incentivized search task and timed Sudoku rounds, recording decisions and performance across successive rounds. At endline, audio recording also captured teams' ideas in their own words. Appendix Figure C.3 shows tablet assessments in girls' and boys' schools, a student receiving the search-task prize, and teams presenting at the Shark Tank finals.

The tablet protocol also incorporated safeguards for data integrity. Module order was randomized across students, so adjacent students encountered different content, and each item appeared on its own screen, which

[10] In Year 1, surveyors typed each team's pitch in English while the students spoke, translating in real time from Telugu, the native language of both the students and the surveyors; records produced this way could not be scored reliably. The Year 1 cycle nonetheless concluded as announced: the typed records were sufficient to shortlist top teams, and shortlisted teams from both arms pitched at an online finals event held by video conference. Because the clarity of a team's record may itself have affected shortlisting, no idea-quality outcome is constructed from the Year 1 cycle.

[11] Behavioral measures of this kind, from grit to negotiation, have been implemented in field experiments using paper-based tasks and in-person games (Alan, Boneva, and Ertac 2019; Ashraf, Bau, Low, and McGinn 2020).

limits copying even among students seated side by side (Singh 2024). Every item required a response before the assessment advanced, which held item-level missingness low. Students could guess or stop the assessment by returning the tablet, but could not skip an item and continue. Responses were recorded directly in the survey application and transmitted to the server, avoiding the read errors common to optically scanned paper answer sheets. The digital format allowed the instruments to be piloted and revised rapidly in out-of-sample schools before each wave: an item that underperformed was revised or dropped within two days, faster than paper-based instruments typically allow.[12]

D. Sample, Balance, and Attrition

At baseline, approximately 5,900 students were enrolled in grade 8 across the 80 study schools, about 74 per school. The baseline assessment was completed by 5,215 of these students, or approximately 88 percent (2,618 in treatment schools and 2,597 in control schools). The grade 10 board examination rosters of the same schools list 5,642 students (2,835 in treatment schools and 2,807 in control schools), who form the study cohort for the administrative outcomes.

Within the baseline-assessed sample, treatment and control students were balanced on observed characteristics (Table 2). Across the composition rows and ten student covariates, no treatment-control difference is significant at conventional levels, and joint tests do not reject balance ($p = 0.48$ for student characteristics, $p = 0.97$ for baseline assessment measures).

To examine attrition, I first compare the baseline assessment roster with grade 10 board examination records, which list every student still enrolled in the school system three academic years later. Baseline-assessed students are linked to a board examination record at statistically indistinguishable rates in the two arms: 77.7 percent of control and 76.8 percent of treatment students ($p = 0.68$).[13]

I next compare attendance at the midline and endline assessments across arms (Appendix Table A.1). Among control students, 82 percent attended midline and 76 percent attended endline, with treatment rates

[12]The assessments were administered using the Qualtrics Offline App, so administration did not require an internet connection. Enumerators worked in teams of five: at each school they registered students by name and unique identifier, distributed pre-charged tablets loaded with the battery in English and Telugu, and supervised the session in a designated classroom. Enumerators received three days of training and followed a standardized script, so instructions were delivered consistently across schools, and they did not assist students with the content of their answers; after each visit the field coordinator checked the synchronized responses for completeness.

[13]A separate linkage check counts each student once within each assessment wave. Of those assessed at baseline, midline, and endline, 20.4, 11.6, and 6.1 percent, respectively, have no linked grade 10 record. These shares do not differ significantly by treatment arm ($p = 0.70$, 0.52, and 0.18, respectively), in regressions with randomization-stratum fixed effects and standard errors clustered by school. Non-linkage may partly reflect school leaving. For context, India's annual dropout rate in grades 9–10 was 16.4 percent in 2022–23, with rates of 19.3 percent among Scheduled Caste students and 23.4 percent among Scheduled Tribe students (Ministry of Education, Government of India 2025).

TABLE 2—EXPERIMENTAL SAMPLE AND BASELINE BALANCE

	Control mean	Treatment mean	Treatment−Control	*p*-value	*N*
Design					
Randomized schools	40	40			80
Students assessed at baseline	2,597	2,618			5,215
Panel A. Student characteristics					
Female students (%)	54.4	51.6	-2.8	0.807	5,215
Tribal-welfare school (%)	48.0	48.2	0.2	0.984	5,215
Rural school (%)	62.1	63.0	0.9	0.937	5,215
Age (years)	13.336	13.285	-0.061	0.169	5,215
Household SES index (SD)	0.054	-0.048	-0.114	0.212	3,566
Household computer literacy (0–1)	0.383	0.403	0.026	0.399	3,566
Joint test *p*-value				0.483	3,566
Panel B. Baseline assessment measures					
Mathematics score (%)	41.6	41.0	-0.4	0.837	5,215
Science score (%)	52.4	52.4	0.1	0.884	5,215
Fluid intelligence (% correct)	57.7	57.9	0.2	0.900	5,215
Emotional intelligence (% correct)	52.0	51.4	-0.5	0.552	5,215
Exploration index (SD)	0.000	-0.021	-0.020	0.591	5,215
Big Five (mean item response)	3.18	3.17	-0.01	0.694	5,215
Risk preference (% of endowment)	63.0	63.4	0.4	0.689	5,210
Joint test *p*-value				0.974	5,210

Notes: This table reports baseline summary statistics and balance. The student sample is students who completed the baseline assessment. Each row reports means by treatment arm; the treatment-control difference is the coefficient on treatment assignment from a regression of the covariate on treatment and randomization strata, with standard errors clustered at the school level; *N* is the number of students with nonmissing values for that covariate. The female, tribal-welfare, and rural rows are the stratification variables and report raw differences without strata controls. All schools are single-gender residential schools run by one of two government societies: social-welfare schools serve predominantly Scheduled Caste students, and tribal-welfare schools serve predominantly Scheduled Tribe students. School characteristics and their balance are reported in Appendix Table A.2. Joint test *p*-values are from Wald tests of all panel covariates in a regression of treatment assignment on the covariates and randomization strata. The household SES index is the first principal component of eight household asset indicators (air conditioner, car, refrigerator, laptop, scooter, washing machine, and father's and mother's mobile phones), standardized. Household computer literacy equals one if a member of the student's household knows how to use a computer; the SES index and household computer literacy are measured at midline and use non-imputed values only. Mathematics, science, fluid intelligence, and emotional intelligence are percent correct on the baseline instruments. The baseline exploration index is the equal-weighted mean of three search behaviors from the round-level baseline exploration game (initiation, discovery, and convergence), each standardized to the control group, with convergence reversed so that higher values indicate more and better-directed search. Because the components are standardized to the control group, the control mean of the index is zero by construction. For brevity, the Big Five row reports the equally weighted mean of the five trait-specific item means on the 1–5 scale, with neuroticism reversed. None of the five traits differs significantly between treatment and control (all individual $p > 0.28$; joint $p = 0.55$). Risk preference is the average share of the endowment bet across nine incentivized rounds, following the investment game of Gneezy and Potters (1997).

0.1 percentage points lower and 4.1 points higher respectively, neither difference significant ($p = 0.93$ and 0.12).[14]

Finally, I examine whether attrition changed the composition of the sample differently across arms. No

[14]These comparisons use school-level regressions of the share of the cohort assessed on treatment and randomization-stratum fixed effects, with heteroskedasticity-robust standard errors.

baseline covariate predicts endline observation differently by arm, individually or jointly ($p = 0.50$). Students not observed at endline do not differ significantly from those observed in their externally graded board scores. Their mathematics scores are 0.04 standard deviations higher and their science scores 0.03 lower. Students assessed at endline are also balanced across arms on all nine baseline covariates. Appendix Table A.3 reports all four tests.

IV. Measurement and Outcomes

The study measured four sets of outcomes using the Innovation-S scale, a rubric constructed for this study and applied to the ideas teams submitted to the end-of-year Innovation Challenge; an incentivized search task that records how students explore alternatives and improve on their solutions; standardized assessments of the remaining innovation-relevant skills, adapted from established batteries; and academic tests together with grade 10 state graduation examination records.

A. The Innovation Challenge and the Innovation-S Scale

The main results examine whether treatment increased participation in the Innovation Challenge and improved the quality of submitted ideas.

Participation is measured by whether a student who completed the individual endline assessment is named on a submitted team roster. Teams submitted their ideas on the same day as the individual assessment and were asked to list all members, including those absent that day.

To measure the quality of submitted ideas, I developed a new instrument for this study, the Innovation-S scale. The scale translates OECD and Eurostat's definition of innovation, a new or improved product made available to its users (OECD and Eurostat 2018), into a rubric for the school setting. Its items were developed from interviews with 14 adult innovators, including a Nobel laureate, social entrepreneurs, and patent holders, and a survey of 44 adults involved in developing products they regard as innovative, including founders, product managers, and executives.[15] Innovation-S assesses the novelty and potential usefulness of ideas.[16]

Teams submitted drawings and audio recordings of their ideas to a surveyor visiting their school. Students in both arms received identical circulars explaining the submission guidelines and shortlisting criteria for the

[15]Text analysis of the interview transcripts using NVivo informed the development of 14 initial items. The Qualtrics survey asked respondents to identify an innovative product or process they knew firsthand and rate statements about its characteristics on a five-point agreement scale. These statements covered novelty, benefits to users, accessibility, and the development process. Respondents also reported users reached, funding, and revenue. Following the survey, seven of the 14 items were retained. Appendix Table C.1 presents the final scale's items, construction, and validation.

[16]Novelty captures whether an idea offers a new solution, improves or combines existing solutions, or transfers a solution to a different context, and is measured as the average of four five-point items. Value-add captures the usefulness of an idea to its intended users and is measured as the average of three five-point items. Appendix Table C.1 presents the novelty and value-add items and their construction and validation. An additional five-point item measures feasibility: whether intended users can access the proposed solution, considering cost and effort.

Shark Tank finals at the beginning of each school year and again shortly before submissions were collected. The submissions described the problem, proposed solution, intended users, and improvement over existing alternatives. Treatment teams were not allowed to submit photographs of prototypes but only a hand-drawn illustration of their idea (see Appendix Figure C.1 for an example), ensuring a common basis for evaluating submissions from both arms.

Two trained evaluators independently evaluated the submitted drawings and transcripts of the audio recordings. Evaluators read the submissions sorted by sector and were not informed of a team's name, school, or treatment status.[17] They scored each idea using the Innovation-S rubric, rated its feasibility, and stated their willingness to fund it, a hypothetical amount between ₹0 and ₹1,000.

Two additional measures support the analysis of idea quality: the direction of ideas and funding received at the Shark Tank finals. The direction of ideas is measured using teams' chosen sectors and intended users. Teams selected one of nine sectors on the submission form. I classify intended users from a separate short written response using keyword rules. Appendix Figure A.6 reports the resulting distributions.

Funding is measured as the total amount paid to a team by professional jurors at the Shark Tank finals in Hyderabad, the state capital. The 40 highest-rated teams from each arm were invited, with travel costs covered. Each team pitched its idea live to a panel typically comprising two jurors. Each juror could allocate up to ₹10,000 to a team at the juror's own discretion. Jurors were not informed of schools' treatment status or that the competition was part of a study.

Appendix C documents the scoring protocol and rater validation to establish the reliability of the ratings, and shows submitted ideas across the quality distribution to illustrate what higher and lower scores represent (Figure C.2).

B. An Incentivized Search Task

The Innovation Challenge measures what teams produced; this task measures how individual students search for solutions. The quality of a submitted idea does not show whether students tried alternatives or used feedback to improve their solution. Such search requires choosing between using a known solution and trying an alternative, the exploration–exploitation trade-off emphasized by March (1991). Exploitation uses a solution with a known payoff; exploration tests an alternative whose payoff is not yet known. Students

[17]The evaluators were research assistants recruited and trained by the data collection agency to apply the Innovation-S rubric. Together, they brought policy research experience and scientific training relevant to assessing proposed solutions and their usefulness to intended users. I had no direct contact with either evaluator, and neither was informed of the experimental design. The two evaluators' index scores correlate at 0.30 across the 1,181 submissions both scored. For comparison, the professional jurors at the finals, scoring the same live pitch, have correlations of 0.41 on Innovation-S and 0.24 on funding. Although evaluator disagreement adds measurement noise, the estimated treatment effect is positive using either evaluator's scores separately (0.225 and 0.179 standard deviations; Appendix Table C.1), with no statistically detectable difference between the two estimates ($p = 0.48$).

practice this search in the curriculum's ideate and make-and-test stages, in which teams generate alternatives, test solutions, and revise them using feedback (Section II.B).

To measure individual search, students in both arms completed the same incentivized task at endline, adapted from the experimentation game of Ederer and Manso (2013). Each student runs a street-food stand using an offline Qualtrics app on a tablet for 12 rounds. Small stands of this kind are common across India, and the curriculum teaches exploration but not business practices, so students in both arms encounter the task with equal familiarity. Each round the student chooses two categorical inputs, a stand type (fresh juice, pani puri, or ice cream) and a location (market, homes, or school), and three quantitative inputs, each in whole units: working hours (6 to 15), the number of menu items (3 to 12), and a selling price (₹11 to 20). Together these give 9,000 configurations.

In the first round, students begin with the same information: a worked example showing a fresh-juice stand configuration that earned ₹30. Beyond the worked example, students learn about profits only through feedback on their own choices. From the second round onward, the screen displays their best configuration and profit so far, while a summary table after each round shows the choices and profits from all completed rounds. Students select all inputs in the first round, while subsequent rounds retain the previous selections, allowing students to repeat the configuration, vary a single input, or vary multiple inputs simultaneously. Disengagement therefore registers as the absence of search rather than as noise. The announced prize, a geometry box, is awarded for the highest hourly profit discovered in the class.

Each stand has an optimal configuration that yields its maximum hourly profit. Formally, profit under stand s in round t is

$$\pi_s(\mathbf{c}_t) = \max\left\{0,\ V_s - \lambda_s \sum_d |c_{td} - c^*_{sd}|\right\}, \tag{1}$$

where V_s is the stand's maximum profit, λ_s its penalty rate, and c_{td} and c^*_{sd} are the student's chosen value and stand s's optimal value for input d, respectively. Stand type determines the payoff parameters; distance is calculated over the other four inputs: working hours, number of menu items, price, and location code. For each input, $|c_{td} - c^*_{sd}|$ measures the absolute difference between the chosen and optimal values, and the sum of these differences is the total distance from the stand's optimum. A one-unit deviation in any input reduces profit by the same amount, λ_s, for that particular stand, subject to the zero-profit floor. Profit is positive if and only if the summed distance is less than V_s/λ_s.

Stands with higher maximum profits impose larger penalties for deviations from their optimal inputs. Fresh juice, the example stand, has the lowest maximum and the smallest penalty ($V = 120$, $\lambda = 10$), pani puri is intermediate in both ($V = 180$, $\lambda = 20$), and ice cream has the highest maximum and the steepest

penalty ($V = 250$, $\lambda = 30$). Higher maximum profits coincide with a higher likelihood of earning nothing: zero profit occurs in 25.7 percent of fresh-juice configurations, 54.0 percent of pani-puri configurations, and 63.3 percent of ice-cream configurations (Appendix Table D.1). Students are not told the stands' maximum profits, optimal inputs, or penalty rates.

The task measures three aspects of exploratory behavior: whether students initiate search, how widely they explore, and whether their search produces better solutions. First, exploration involves leaving a known, profitable solution to try an alternative with an unknown payoff. Initiation is measured by whether students choose a stand other than the example stand in round 1. This choice precedes feedback on their own decisions, so all students face it with the same task information.

Second, the task encourages continued exploration by rewarding the best solution the student discovers.[18] Once a student has found a profitable configuration, an unsuccessful attempt cannot reduce the best profit already achieved. The objective therefore tolerates the early failure emphasized by Manso (2011). It also gives the student reason to try unfamiliar alternatives rather than repeat a known configuration that cannot improve the recorded maximum. Breadth is the number of distinct stand-and-location combinations tried across the 12 rounds. I use these categorical choices because small adjustments to hours, menu items, or price can refine an existing alternative without broadening the search.

Third, I measure search success using Discovery and Convergence. Discovery counts the rounds in which profit exceeds the student's previous best. Convergence is the minimum distance achieved across the twelve rounds, with each round's distance measured from the optimum for the stand used in that round. The displayed personal best gives students a benchmark against which to assess each new attempt.

The exploration index combines four aspects of search: whether the student departs from the example stand, how widely they range, how often they improve on their own best, and how close they come to the optimum. It is the equal-weighted mean of Initiation, Breadth, Discovery, and Convergence, with each component standardized to the control group and Convergence entering with reversed sign.

Appendix D reports the payoff parameters and optimal inputs for each stand.

C. The Remaining Skill Tasks

I assessed emotional intelligence, fluid intelligence, creativity, and structured problem solving using separate instruments. Table 3 pairs each skill with its instrument. These assessments draw on established

[18]The baseline version of the task followed Ederer and Manso (2013): students ran a single simulated lemonade stand and chose its location, ingredients, and price. Its cumulative-earnings objective proved hard for students of this age to follow; the endline task instead rewarded the best round, a goal a grade 9 student in this context could pursue. The best-round objective also aligns incentives with discovering better solutions: under cumulative earnings, repeating a known profitable configuration increases the total, whereas here it cannot improve the recorded maximum. Performance on the baseline task enters the endline specifications as a pre-treatment control.

batteries and published experimental designs, adapted to the setting and piloted with out-of-sample students. Appendix D reports their adaptation, scoring, and reliability. Item-based scores use a two-parameter item-response model and are standardized to the control group.

The find and learn stages practice perspective-taking and pattern recognition. I assess these skills through tests of emotional and fluid intelligence. Emotional intelligence is measured with the Reading the Mind in the Eyes Test (RMET) (Baron-Cohen et al. 2001), which shows the student a photograph of the eye region of a face and asks which emotion best describes what the person is feeling or thinking. Weidmann and Deming (2021) and Alan and Kubilay (2025) have also used the RMET to measure social and emotional intelligence. Fluid intelligence is measured with matrix-reasoning items from the International Cognitive Ability Resource, a public-domain battery validated against standard intelligence tests (Condon and Revelle 2014). Each item presents a grid of geometric patterns with one cell missing, which the student completes. Appendix Figures D.1 and D.2 show a sample item from each test.

The ideate stage practices divergent thinking and evaluative judgment. Three written tasks adapted from the PISA creative-thinking assessment, first administered in 2022, measure idea generation, idea evaluation, and written expression (OECD 2024). In the idea-generation task, students propose three improvements to a standard bicycle. The rubric distinguishes responses that merely name bicycle parts from those proposing standard features, common improvements, or uncommon and innovative additions. Idea evaluation asks the student to allocate funds between two innovations and justify the allocation. The justification is scored according to whether it addresses cost, usefulness, or user needs rather than merely stating a preference. In the written-expression task, students write a short story about an image. Responses are scored for narrative structure and use of the image's elements. Responses are open-ended text in English or Telugu, scored against task-specific rubrics by a large language model. A research assistant blind to treatment status verified all assigned scores and confirmed more than 98 percent without changes. Appendix Figure D.3 shows the task prompts and stimulus images and Appendix Table D.3 reports each task's rubric with example responses at every score level.

The make-and-test stage practices structured problem solving. I assess this skill with a five-round Sudoku task using four-by-four puzzles. A Sudoku puzzle is a well-defined problem: the rules and goal are specified, and each step can be checked against the rules (Newell and Simon 1972). The task requires little subject-specific knowledge, and the reduced grid makes it accessible to students across the range of academic attainment.

Before each round, students choose between an easy and a hard Sudoku puzzle. The measure adapts the incentivized easy-versus-hard choice design of Alan, Boneva, and Ertac (2019), with a Sudoku puzzle in

TABLE 3—CURRICULUM SKILLS, PRACTICE, AND STANDARDIZED MEASURES

Skill	Curriculum location	How the curriculum practiced it	Standardized measure (source)
Creative confidence	Ideate; Make and test; independent design	Generated multiple solutions, attempted solutions before seeing examples, built mock-ups, revised after feedback, and progressed from guided tasks to open-ended problem discovery.	Not separately assessed
Critical thinking	Find; Learn; Ideate; critique and selection	Distinguished problems from non-problems, investigated causes and stakeholders, examined design changes, compared alternatives, and evaluated ideas against explicit criteria.	Fluid intelligence, ICAR (Condon and Revelle 2014); idea evaluation, PISA (OECD 2024)
Creativity	Ideate; redesign challenges; sensor applications	Brainstormed under time and material constraints, redesigned objects, combined objects and features, and developed uses for sensors and outputs.	Idea generation and written expression, PISA (OECD 2024)
Empathy	Find; user-focused design; testing	Observed and interviewed users, identified their problems, designed for specified users in semi-open challenges, and gathered feedback on prototypes.	Reading the Mind in the Eyes Test (Baron-Cohen et al. 2001)
Problem solving	Full five-stage cycle	Defined and investigated problems, generated and selected solutions, planned resources, and built, tested, and revised prototypes with increasing independence.	4×4 Sudoku, difficulty-weighted (Alan, Boneva, and Ertac 2019); incentivized search task (Ederer and Manso 2013)
Collaboration	Team work; co-team feedback; peer review	Worked in fixed teams, divided tasks, compared ideas, exchanged feedback, and participated in class and showcase discussions, facilitated by student leaders.	Not separately assessed
Presentation	Present; Pause; Exhibition/showcase	Presented problems, evidence, solutions, prototypes, and feedback to class and school audiences using presentation cards and reflection prompts.	Recorded pitch (Section IV.A)
Communication	Interviews; peer feedback; presentations	Questioned users, articulated problems and design choices, listened to alternative perspectives, and explained prototypes and revisions.	Recorded pitch (Section IV.A); grade 10 board-examination language scores (Section IV.D)

Notes: The first three columns describe the curriculum's intended skills and practices, as mapped by its developer, Inqui-Lab. The final column lists the assessments selected by the research team, drawn from psychology (RMET and the ICAR fluid intelligence test), education (PISA creative-thinking tasks), and economics (Sudoku and the incentivized search task). Grade 10 language scores provide an indirect measure of communication. The curriculum mapping draws on Inqui-Lab's product note, framework, session plans, student workbooks, portfolios, student-leader handbook, and Year-1 observation materials.

place of their number-matching real-effort task. Students receive a larger prize for solving a hard puzzle than an easy one, and no prize for an unsolved puzzle. Choosing the hard puzzle entails a trade-off between a larger prize and a higher probability of failure. In the first round, half of the students, selected at random, played the hard puzzle regardless of their choice, and from the second round onward every student played the puzzle they chose. Problem solving is the average score across the five rounds, with hard-puzzle scores multiplied by 1.5 before averaging. The choices themselves measure challenge-seeking, defined as the share of rounds 2–5 in which the student chose the harder puzzle. Students receive no performance feedback between rounds, so this measure captures willingness to attempt a harder problem rather than persistence following failure.

D. Academic Outcomes

To measure academic achievement, the study uses baseline and endline tests and grade 10 state graduation examination scores. The baseline academic score pools responses to 10 mathematics items and 17 science

items in a single two-parameter item-response model. At endline, students completed 22 mathematics items and 20 science items on the tablet, alongside the skill tasks. I developed both tests using grade 8 and grade 9 state mathematics and science textbooks. The mathematics test also incorporated a few items adapted from the endline assessments used by Muralidharan, Singh, and Ganimian (2019). Both tests were piloted in two out-of-sample schools.

At endline, students also reported their time use and school attendance. Students reported the hours spent in a typical school week on school periods, homework, sports, the Innovation Challenge, and other activities, which I express as shares of their total reported time. School attendance is the share of the preceding twelve school days on which the student reported being present.

The grade 10 board examinations took place one year after the program ended. Outcomes are the scores in the six subjects (mathematics, science, social studies, and three language papers), the official total, and an indicator for passing, with scores standardized to the control group.

V. Empirical Strategy

I estimate intent-to-treat effects of the program by ordinary least squares, comparing outcomes across randomly assigned treatment and control schools and conditioning on randomization-stratum indicators and baseline covariates. For outcome Y_{is} of student i in school s, the specification is

$$Y_{is} = \alpha + \beta\, T_s + R_s'\mu + X_i'\gamma + W_s'\theta + \rho B_i + \delta M_i + \varepsilon_{is}, \tag{2}$$

where T_s indicates a treatment school and R_s contains indicators for the three characteristics defining the randomization strata: tribal-welfare management, girls-only composition, and rural location. The vector X_i contains the student's age, household SES index, and household computer literacy, with indicators for imputed values; W_s contains the pre-treatment school characteristics listed in Appendix Table A.2. The variable B_i is the student's baseline measure for the corresponding outcome family, and M_i indicates students with a missing baseline measure; missing baseline scores are imputed at the school-section mean, so these students remain in the estimation sample. The coefficient β is the intent-to-treat effect of assignment to the curriculum. Standard errors are clustered at the school level.

For team-level idea ratings, I estimate

$$Y_{gs} = \alpha + \beta\, T_s + R_s'\mu + W_s'\theta + X_g'\phi + \varepsilon_{gs}, \tag{3}$$

where Y_{gs} is the rating of team g in school s. The vector X_g contains team averages of members' baseline

academic achievement, exploration index, fluid intelligence, emotional intelligence, household SES index, computer literacy, and age. It also contains three count variables: team size, the number of members with observed baseline scores, and the number without a matched individual record. Standard errors are clustered at the school level.

For submission-adjusted idea ratings, I randomly group non-submitters into teams of up to four students within each school and assign these teams a raw score of zero before standardization. Holding submitted teams fixed, I repeat the grouping and estimation 1,000 times and report the average coefficient. The reported variance combines the average school-clustered variance with the variance of coefficients across groupings.

Appendix G lists deviations from the registered pre-analysis plan.

VI. Results

Participation in the Innovation Challenge. Treatment increased participation in an innovation competition offered on the same terms to students in both arms (Table 4, row 1). Among students assessed on submission day, participation rose by 8.0 percentage points from a control mean of 73.3 percent, significant at the 1 percent level. This corresponds to a roughly 30 percent reduction in nonparticipation.

Because submission-day attendance is a post-treatment outcome, I also estimate participation using the baseline-assessed roster, a denominator fixed before treatment. Treatment increased participation in this sample by 5.5 percentage points from a control mean of 56.4 percent ($p = 0.031$).[19]

Idea Quality. Treatment increased submission-adjusted ratings of novelty by 0.187 control-group standard deviations, value-add by 0.146, feasibility by 0.109, and willingness to fund by 0.184 (Table 4, Panel B). These estimates include non-submitters assigned zero scores before standardization. To examine how the gains vary across rating thresholds, Appendix Figure A.1 compares the share of teams whose ideas meet or exceed each threshold. At every quality threshold from the no-submission margin through the middle of the distribution, the share of teams clearing the threshold is higher in treatment schools, with the gap narrowing only in the far upper tail.

Restricting the comparison to teams that submitted an idea, treatment-school submissions also received higher ratings (Table 4, Panel C). The estimates are 0.265 standard deviations for novelty, 0.202 for value-add, and 0.206 for willingness to fund, while the feasibility estimate is 0.100 and is not statistically significant.

[19] Students named on submitted team rosters count as participants regardless of submission-day attendance, while students who left the school system remain in the denominator as nonparticipants. Both estimates use submitted team rosters to identify participation (Section IV.A). Presence at the endline assessment did not detectably differ by arm (a difference of 2.4 percentage points, $p = 0.29$); Appendix Table A.3 reports coverage at every stage and the differential-attrition tests.

TABLE 4—EFFECTS ON INNOVATION CHALLENGE PARTICIPATION AND IDEA RATINGS

Outcome	Treatment effect	SE	Sharpened q	N
Panel A. Participation				
Participated in the Innovation Challenge (pp)	8.0***	(2.1)	[0.002]	4,388
Panel B. Submission-adjusted idea ratings				
Innovation-S novelty subscale (SD)	0.187***	(0.054)	[0.002]	1,493
Innovation-S value-add subscale (SD)	0.146***	(0.052)	[0.005]	1,493
Feasibility rating (SD)	0.109**	(0.047)	[0.009]	1,493
Evaluator willingness to fund (SD)	0.184***	(0.062)	[0.004]	1,493
Panel C. Idea ratings among submitted teams				
Innovation-S novelty subscale (SD)	0.265***	(0.070)	[0.002]	1,225
Innovation-S value-add subscale (SD)	0.202***	(0.069)	[0.007]	1,225
Feasibility rating (SD)	0.100	(0.062)	[0.029]	1,225
Evaluator willingness to fund (SD)	0.206***	(0.073)	[0.007]	1,225

Notes: Panel A reports participation among 4,388 students assessed at endline. Participation equals one if a student is named on a submitted team roster; the effect is in percentage points (control mean: 73.3 percent). Panel B reports submission-adjusted idea ratings for 1,493 team units: 1,216 scored submitting teams with at least one endline-assessed member and 277 constructed teams containing the 1,002 remaining students. Constructed teams contain up to four students from the same school. Non-submission receives a raw score of zero on each rating, including ₹0 for willingness to fund, before standardization. Rating effects in Panel B are in standard deviations of the control group in this augmented sample. Panel C reports the same four ratings for all 1,225 scored submitting teams, including nine teams with no endline-assessed member, in standard deviations of the control group among submitted teams. Because treatment affects submission, the Panel C estimates are conditional on submission.

Student and team controls follow equations (2) and (3); participation uses baseline mathematics and science as its baseline measures. Panel B estimates average over 1,000 random groupings of non-submitters; their variances combine the average school-clustered variance and the between-grouping variance. Participation and Panel C standard errors are clustered by school. Stars test each effect against zero using the clustered p-value. Bracketed values are Anderson sharpened false-discovery-rate q-values, computed within the five outcomes of Panels A and B and, separately, within the four outcomes of Panel C. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Because treatment increased submission, the comparison among submitting teams may reflect differences in who participated as well as differences in idea quality. Submitting teams do not differ significantly in their average baseline characteristics. Estimated lower Lee bounds accounting for differential submission are positive for novelty, value-add, and willingness to fund, but not feasibility (Appendix Table A.4). These bounds do not incorporate sampling uncertainty.

Jury Funding at the Shark Tank Finals. The Shark Tank finals provide a second assessment of the ideas through actual jury funding. The 40 highest-rated teams from each arm were invited to participate. The jurors also scored each pitch on the same Innovation-S rubric before allocating funding, and their rubric scores were strongly correlated with their funding allocations (within-juror $r = 0.77$). Figure 3 shows individual funding allocations and differences between treatment and control funding quantiles.

Among these finalists, treatment teams raised ₹1,709 more on average than control teams, a difference of 27 percent, but neither the mean difference (randomization inference $p = 0.28$) nor a rank-based comparison of the full distributions (rank-sum $p = 0.42$) is statistically significant. The exploratory quantile comparison in

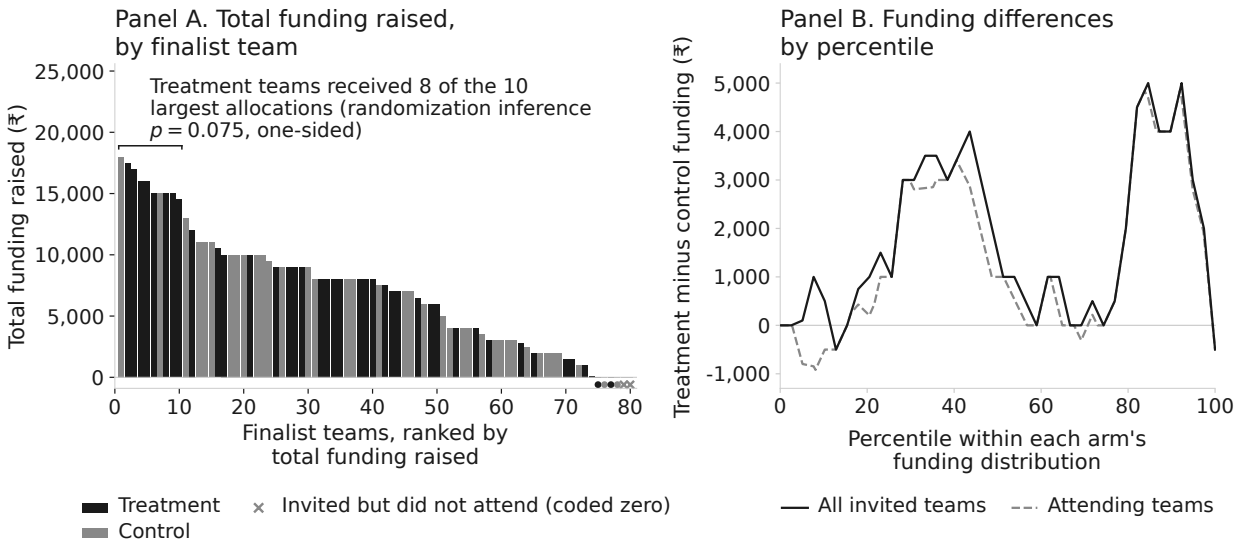


FIGURE 3. JURY FUNDING AT THE SHARK TANK FINALS

Notes: The sample is the 80 finalist teams of the Innovation Challenge, the top 40 teams from each arm ranked on the evaluators' ratings, invited to the Shark Tank-style finals. The outcome is the total funding a team raised from its jurors, in rupees. Panel A ranks all 80 teams by total funding raised; the two invited control teams that did not attend are coded as raising zero and marked with crosses. Panel B plots the treatment-control difference in funding at corresponding percentiles of each arm's distribution. The solid line includes all 80 invited teams; the dashed line excludes the two control no-shows and recomputes the quantiles among the 78 attending teams. Quantiles use linear interpolation; the endpoints are the sample minimum and maximum. The quantile comparisons are descriptive, and no confidence intervals are shown. The randomization-inference p-value in Panel A is obtained by permuting treatment assignment at the school level within randomization strata (5,000 draws) among the 78 attending teams, whose schools are observed. Funding estimates appear in Appendix Table A.5, and funding survival curves in Appendix Figure A.2.

Panel B shows positive funding differences in the lower-middle and upper portions of the finalist distribution, with smaller differences in between. These positive differences persist when the two control no-shows are excluded. Of the 10 teams that raised the most money, 8 were treatment teams (one-sided randomization inference $p = 0.075$), though the single highest-funded team was a control team.

Because finalists were selected using post-treatment ratings, treatment may affect which teams reach the finals as well as the quality of their ideas. The funding comparison therefore concerns each arm's selected finalists, rather than the effect of treatment on a fixed set of teams. The funding estimates are positive but imprecise, so I report this comparison as directional evidence.

Appendix Table A.5 reports these estimates together with a robustness check comparing teams only within jury panels. Appendix Figure C.3 shows teams presenting to the jury panels.

Innovation-Relevant Skills. Treatment improved students' performance on the incentivized search task (Table 5, Panel A). The exploration index rose by 0.094 index points. Initiation, the probability of leaving the example stand in the first round, rose by 5.6 percentage points. Breadth, the number of distinct stand-and-location combinations tried, increased by 3.6 percent relative to the control mean.

TABLE 5—EFFECTS ON INNOVATION-RELEVANT SKILLS

Outcome	Treatment effect	SE	Sharpened q	N
Panel A. Search behavior in the incentivized search task				
Exploration index (points)	0.094***	(0.025)	[0.002]	4,388
Initiation: left the example stand, round 1 (pp)	5.6***	(1.8)	[0.006]	4,388
Breadth: unique stand-location combinations tried	0.161*	(0.092)	[0.035]	4,388
Discovery: rounds improving on the personal best	0.101**	(0.039)	[0.012]	4,388
Convergence: minimum distance from the optimum	-0.256**	(0.115)	[0.020]	4,388
Panel B. Creativity				
Idea generation (SD)	0.314***	(0.078)	[0.001]	4,388
Idea evaluation (SD)	0.121*	(0.066)	[0.075]	4,388
Written expression (SD)	0.040	(0.065)	[0.219]	4,388
Panel C. Analytical reasoning				
Fluid intelligence (SD)	0.111**	(0.050)	[0.032]	4,388
Problem solving (SD)	0.142**	(0.064)	[0.032]	4,388
Panel D. Emotional intelligence and challenge-seeking				
Emotional intelligence (SD)	0.031	(0.048)	[1.000]	4,388
Challenge-seeking (SD)	0.019	(0.078)	[1.000]	4,388

Notes: The table reports treatment effects on innovation-relevant skills (Sections IV.B and IV.C), estimated at the student level on the 4,388 students who completed the endline assessment. Initiation is reported in percentage points; Breadth, Discovery, and Convergence are in natural units. The exploration index is the equal-weighted mean of the four control-standardized components, with Convergence reverse-signed. Outcomes in Panels B–D are standardized to have control-group mean zero and standard deviation one. **Panel A** reports search behavior in the incentivized search task; the outcomes are the four margins of a search and their exploration index. Initiation equals one if the student's round-1 stand differs from the example stand (control mean 57.9 percent); Breadth counts distinct stand-and-location combinations (control mean 4.50 of 9); Discovery counts rounds improving on the personal best (control mean 1.26); Convergence is the minimum distance achieved across the twelve rounds, with each round's distance measured from the optimum for the stand used in that round (control mean 6.30). **Panel B** reports three PISA-adapted written tasks scored by rubric: idea generation (rubric ratings of three proposed bicycle improvements), idea evaluation (the justification of a budget allocation between two innovations), and written expression (a short story). **Panel C.** Fluid intelligence is a two-parameter item-response score from twelve ICAR items. Problem solving is the standardized difficulty-weighted mean score across five Sudoku rounds. **Panel D** reports emotional intelligence, from the eleven-item Reading the Mind in the Eyes Test (two-parameter item-response model), and challenge-seeking, the share of rounds 2–5 in which the student chose the harder puzzle in the Sudoku task.

Every regression controls for randomization strata, school infrastructure, and individual covariates (age, socioeconomic index, computer skills), together with the panel's pre-treatment baseline measure and a missing-value indicator: the baseline exploration index for Panels A and B, baseline fluid intelligence for Panel C, and, in Panel D, the baseline eyes test for emotional intelligence and baseline conscientiousness for challenge-seeking. Standard errors, in parentheses, are clustered at the school level, the level of treatment assignment; significance stars refer to the clustered p-value. Values in square brackets are Anderson sharpened false-discovery-rate q-values, computed within each panel family. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Search also produced better solutions. Discovery, the number of rounds improving on the student's previous best profit, increased by 8 percent relative to the control mean. Convergence, the minimum distance achieved across the twelve rounds, was 4.1 percent lower than the control mean. Each round's distance is measured from the optimum for the stand used in that round. The index and all four components have sharpened false-discovery-rate q-values below 0.05.

The largest standardized skill gain was in idea generation on an individual task. This score, based on rubric ratings of three proposed improvements to a bicycle, rose by 0.314 standard deviations (Panel B).

Performance on this task may also reflect students' application of knowledge acquired in the program. The estimated effects on the other two tasks were smaller. Idea evaluation rose 0.121 standard deviations; the effect on written expression is 0.040 and statistically indistinguishable from zero.

Gains also extended to fluid intelligence and problem solving (Panel C). Fluid intelligence rose 0.111 standard deviations, and the difficulty-weighted Sudoku score rose 0.142 standard deviations. Both effects survive adjustment for multiple hypotheses. The problem-solving gain is similar under alternative scoring methods that remove the difficulty weighting or adjust for difficulty choice (Appendix Table A.6).

Treatment had no detectable effects on emotional intelligence or challenge-seeking (Panel D). The estimated effects are 0.031 and 0.019 standard deviations, respectively.

I also examine whether students' individually assessed skills are associated with the quality of the ideas they produce in teams. Teams with higher average scores on the individual skill assessments tended to submit higher-rated ideas (Appendix Figure A.3). The estimated relationships are positive for all four skills in both treatment arms.

Academic Outcomes. The program produced no detectable declines in academic achievement at endline or on the grade 10 board examinations (Table 6). Sessions were held during after-school study or other non-academic periods, and the curriculum did not teach examination subjects. I therefore examine whether time devoted to the program came at a cost to academic achievement. Panel A reports the endline academic assessment. The effect on mathematics is -0.007 control-group standard deviations and the effect on science is 0.027; both are indistinguishable from zero, and the 95 percent confidence intervals rule out losses larger than 0.14 standard deviations in mathematics and 0.10 in science.[20]

The time-use evidence suggests that students accommodated the program through changes in non-academic activities, without a detectable reduction in academic engagement. Neither the reported time shares devoted to homework, study, and school periods nor recent school attendance showed a statistically detectable decline (Table 6, Panel A; Appendix Table E.1). Reported weekly time shares were 1.3 percentage points higher for preparation for the Innovation Challenge and also increased for sports. These increases were accompanied by a lower share devoted to other activities, including dance, singing, drawing, and reading (Appendix Table E.1).

Academic achievement showed no detectable average decline on the grade 10 board examinations administered in March 2025, one year after the program ended (Table 6, Panel B). Point estimates are positive

[20] A midline assessment in February 2023, near the end of Year 1 and in the first full academic year after COVID-related school closures, showed lower mathematics scores in treatment schools that did not persist in later waves. Mathematics was 0.137 standard deviations lower ($p = 0.007$), science was unchanged (0.000), and the exploration index was 0.145 standard deviations higher ($p = 0.006$; $N = 4{,}620$, Appendix Table E.2). Appendix F compiles students' open-ended feedback from the same assessment.

TABLE 6—EFFECTS ON ACADEMIC OUTCOMES

Outcome	Treatment effect	SE	Sharpened q	N
Panel A. Endline assessment				
Mathematics	-0.007	(0.067)	[1.000]	4,388
Science	0.027	(0.061)	[1.000]	4,388
Time on homework (% of reported time)	-0.36	(0.45)	[1.000]	4,388
School attendance, past two weeks (%)	1.44	(1.00)	[1.000]	4,388
Panel B. Grade 10 board examinations				
Mathematics	0.112	(0.147)	[0.598]	5,632
Science	-0.006	(0.119)	[0.945]	5,632
Social studies	0.023	(0.093)	[0.945]	5,632
First language	0.149*	(0.088)	[0.338]	5,632
Second language	0.199*	(0.106)	[0.338]	5,631
Third language	0.153*	(0.086)	[0.338]	5,631
Total score	0.135	(0.115)	[0.436]	5,633
Passed (pp)	0.64	(0.83)	[0.598]	5,642

Notes: The table reports treatment effects on academic outcomes at two horizons. **Panel A** reports the endline academic assessment, administered in the same battery as the skill tasks in Table 5, for the 4,388 students who completed the endline. Mathematics and science are item-response scores standardized to the control group. Time on homework is the share of the student's reported weekly time, in percentage points (control mean 19.2 percent), from a five-activity weekly hours grid; grids totaling 80 or more hours per week (28.3 percent of students) had all five activity shares replaced by the rounded mean shares of unflagged classmates in the same school section, and this replacement flag is balanced across arms ($p = 0.73$). School attendance is the percent of the twelve school days in the two weeks before the assessment on which the student was present, recoded from the item asking how many days the student was absent from school (control mean 89.9 percent). **Panel B** reports the March 2025 grade 10 board examinations, whose administrative records cover the entire 5,642-student study cohort. Subject scores are the externally graded examination component, standardized to the control group; the total score is the standardized official total across subjects, which includes internal marks; Passed is an indicator scaled to percentage points (control mean 96.7 percent). Every regression controls for randomization strata, school infrastructure, and individual covariates (age, socioeconomic index, computer skills), together with the row's pre-treatment baseline measure and its missing-value indicator: baseline mathematics for the mathematics rows, baseline science for the science rows, and both baseline scores for the remaining rows. Standard errors are clustered at the school level. Values in square brackets are Anderson sharpened false-discovery-rate q-values, computed within each panel.

in five of the six subjects. Gains in the three language subjects provide suggestive, indirect evidence of improved communication skills, although none survives the false-discovery-rate adjustment.

Grade 10 board scores are concentrated near the maximum, limiting how clearly they distinguish achievement among higher-scoring students. In both mathematics and science, 40 percent of control students scored at least 80 percent of the maximum possible marks. The correlations between board scores and the endline assessment are only 0.20 in mathematics and 0.30 in science. Students scoring at least 80 percent on the board examinations are represented in all five quintiles of the baseline and endline assessments, indicating that high board scores can mask substantial differences in prior academic achievement (Appendix Figure E.1). Related evidence comes from Singh (2024), who documents inflated scores in teacher-administered grade 4 assessments in Prakasam district, in neighboring Andhra Pradesh.

Relationship between Academic Scores and Innovation-Relevant Skills. Academic scores, including the grade 10 board scores that screen entry into science streams, weakly predict which students are strongest in the innovation-relevant skills. Figure 1 plots the four endline skill measures against baseline and endline academic scores and the grade 10 board-examination total among control students. The baseline academic score explains at most 12 percent of the variation in idea generation, problem solving, fluid intelligence, or exploration, and the grade 10 board score at most 10 percent; the endline academic score explains at most 23 percent. A screen limited to the top baseline or grade 10 academic quintile would miss roughly 60 to 80 percent of the students in the top quintile of each skill.

Academic measures correlate 0.35 to 0.50 with one another; their correlations with the four skill measures range from 0.06 to 0.44 (Appendix Figure A.4). These comparisons are descriptive; they are not treatment effects.

Heterogeneity. I compare control-group means and treatment effects across four pre-treatment characteristics: school gender composition, baseline academic score, management, and location (Figure 4). The four characteristics are not independent, as students in girls' schools, social-welfare schools, and urban schools enter the study with higher baseline academic scores, by 0.36, 0.41, and 0.29 standard deviations respectively. A difference in treatment effects by gender composition, management, or location therefore partly reflects a difference by baseline score.

Figure 4, Panel A shows that the curriculum increased participation and idea quality among groups with lower prior academic achievement: boys' schools, students and teams below the baseline academic median, tribal-welfare schools, and rural schools. Benefits also extended to academically advantaged groups. Above-median students increased their participation, and both above-median teams and teams in girls' schools produced more novel ideas. Estimates for social-welfare and urban schools were generally smaller and imprecise. Although estimated gains consistently favored the academically disadvantaged group, most between-group differences were not statistically significant.

Panel B shows that the benefits for students with lower prior academic achievement extend beyond participation and idea quality to innovation-relevant skills. Exploration improved broadly across subgroups, and idea-generation gains reached students in both halves of the baseline academic distribution. Problem-solving gains were also evident among below-median students and in boys' and rural schools, demonstrating that skill development extended to groups entering the study with lower academic scores. This broad reach coexisted with larger gains among academically advantaged groups in some domains. Idea-generation gains were larger among above-median students and in girls' and social-welfare schools; girls' schools gained

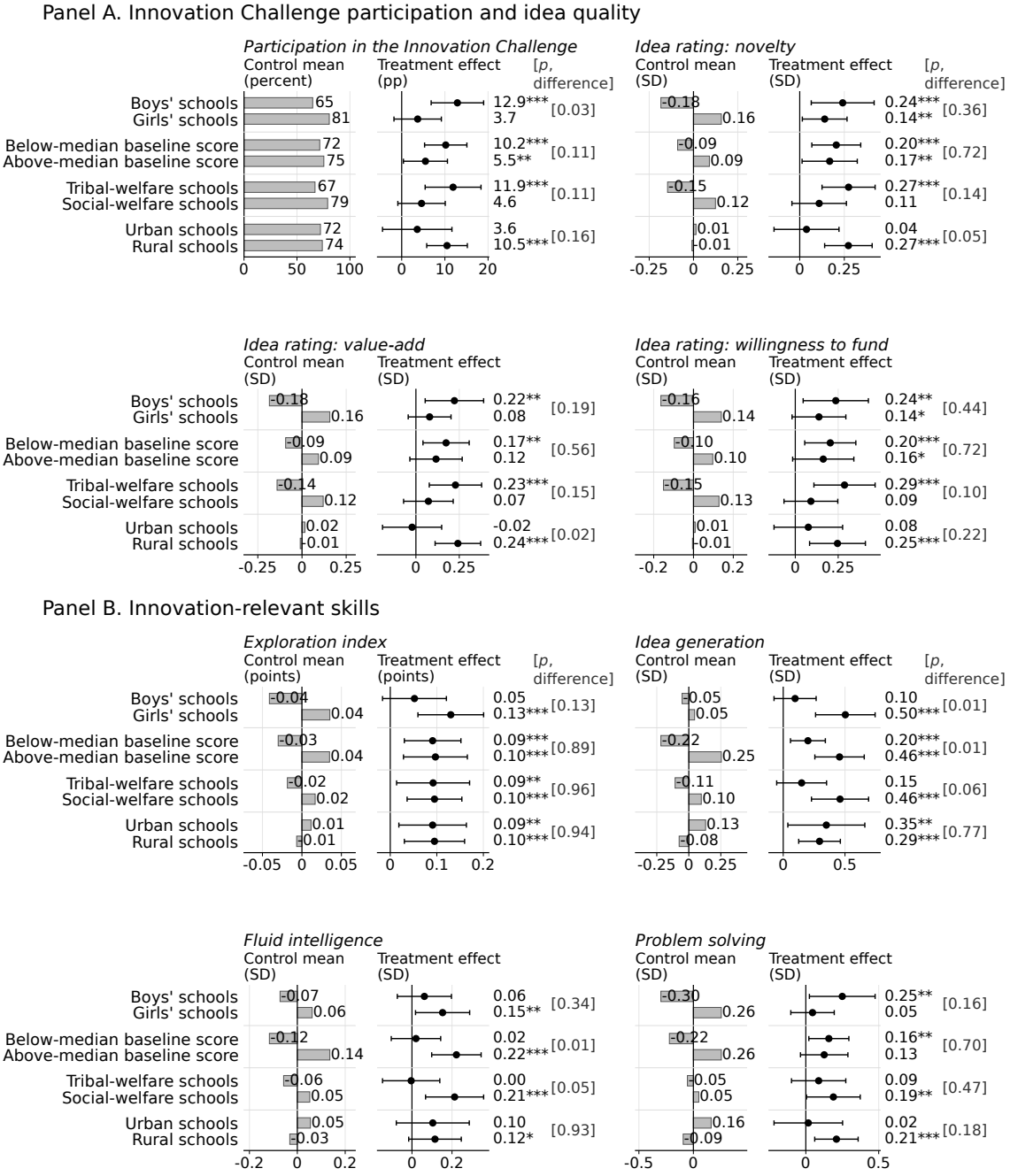


FIGURE 4. TREATMENT EFFECTS BY SUBGROUP

Notes: Bars show control-group means; points show treatment effects and 95 percent confidence intervals for eight subgroups defined by school gender composition, baseline academic score, management, and location. The estimate is printed to the right of the interval, and the value in brackets is the p-value for equality of the two effects in the pair. Participation and the measures in Panel B use the 4,388 students assessed at endline. The idea ratings use 1,493 teams, with teams that did not submit an idea assigned the value of no idea. Idea ratings and skill measures other than the exploration index are in standard deviations of the control group. Baseline-score subgroups are defined by the control-group median of the baseline academic score for students and, for teams, of the team-average score among control teams, recomputed within each grouping of non-submitters. Effects are estimated from regressions that interact treatment with the subgroup indicator and include randomization strata, school infrastructure, individual or team controls, and the baseline measure of the outcome. Standard errors are clustered at the school level and, for idea ratings, incorporate between-grouping variation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

0.50 standard deviations, compared with 0.10 in boys' schools.[21] Fluid-intelligence gains also favored above-median students and social-welfare schools, whereas exploration showed no detectable subgroup differences.[22] The curriculum thus developed skills among students with lower prior achievement while also allowing academically advantaged students to make larger gains on some tasks.

Figure 5 divides baseline academic scores into quintiles to examine where gains occur within the distribution, including whether they reach the lowest-achieving students. Idea generation improved in every quintile, and problem-solving gains reached both ends of the distribution, including a gain of 0.33 standard deviations in the lowest quintile. The average skill gains therefore include meaningful improvements among students with the lowest baseline academic scores. Larger estimated participation gains among lower-achieving students and broadly positive exploration effects without a clear academic gradient also reinforce the earlier subgroup findings.

At the same time, students in the top quintile of baseline academic achievement made the largest estimated gains in idea generation, idea evaluation, fluid intelligence, and endline mathematics. This pattern reinforces the binary evidence that some skill gains favor academically stronger students and shows that their gains also extend to academic performance. For this group, idea generation and endline mathematics increased by 0.63 and 0.23 standard deviations, respectively.[23]

The academic effects became more favorable between midline and endline, particularly among the highest-achieving students. Following the same students across both assessments, the average treatment effect increased significantly, with a significantly larger increase among students in the top baseline quintile than among the remaining students (Panel B). Nevertheless, in the full endline sample, estimates remained negative in the second and middle quintiles, with a middle-quintile decline of 0.15 standard deviations significant at the 10 percent level.[24] The near-zero average endline effect therefore combines academic gains for some students with possible costs for others. Grade 10 mathematics estimates remained imprecise, with no comparable

[21] All schools are single-gender, so these comparisons also reflect differences between boys' and girls' school environments. When gender and baseline-score interactions are included jointly, both remain significant for idea generation, while only the baseline-score interaction remains significant for fluid intelligence. The idea-generation differential is concentrated in girls' schools within the social-welfare society (Appendix Table A.9).

[22] Adjusting for the sixteen interaction tests within each subgroup dimension yields sharpened q-values of 0.079 for the baseline-score differentials in idea generation and fluid intelligence, and 0.210 for the gender differential in idea generation.

[23] The top-quintile estimates are similar when the analysis excludes students with imputed baseline scores: 0.62 standard deviations for idea generation, 0.29 for idea evaluation, 0.26 for fluid intelligence, and 0.22 for endline mathematics.

[24] The binary comparisons show a similar pattern: endline mathematics effects are -0.11 standard deviations below the baseline academic median and 0.12 above it, and -0.18 in tribal-welfare schools and 0.14 in social-welfare schools. Only the tribal-welfare decline is individually significant at the 5 percent level, but both between-group differences are significant.

Among the same 3,641 students assessed at both waves, with quintile membership held fixed, the average treatment effect increased by 0.15 standard deviations ($p = 0.023$). The increase was 0.32 standard deviations larger in the top quintile than among the remaining students ($p = 0.002$). The middle-quintile estimates are -0.22 standard deviations at midline and -0.11 at endline; the increase of 0.12 is not statistically significant ($p = 0.262$). Tests of changes use student-level differences between waves, the same controls, and school-clustered standard errors. Scores retain their wave-specific control-group standard deviations.

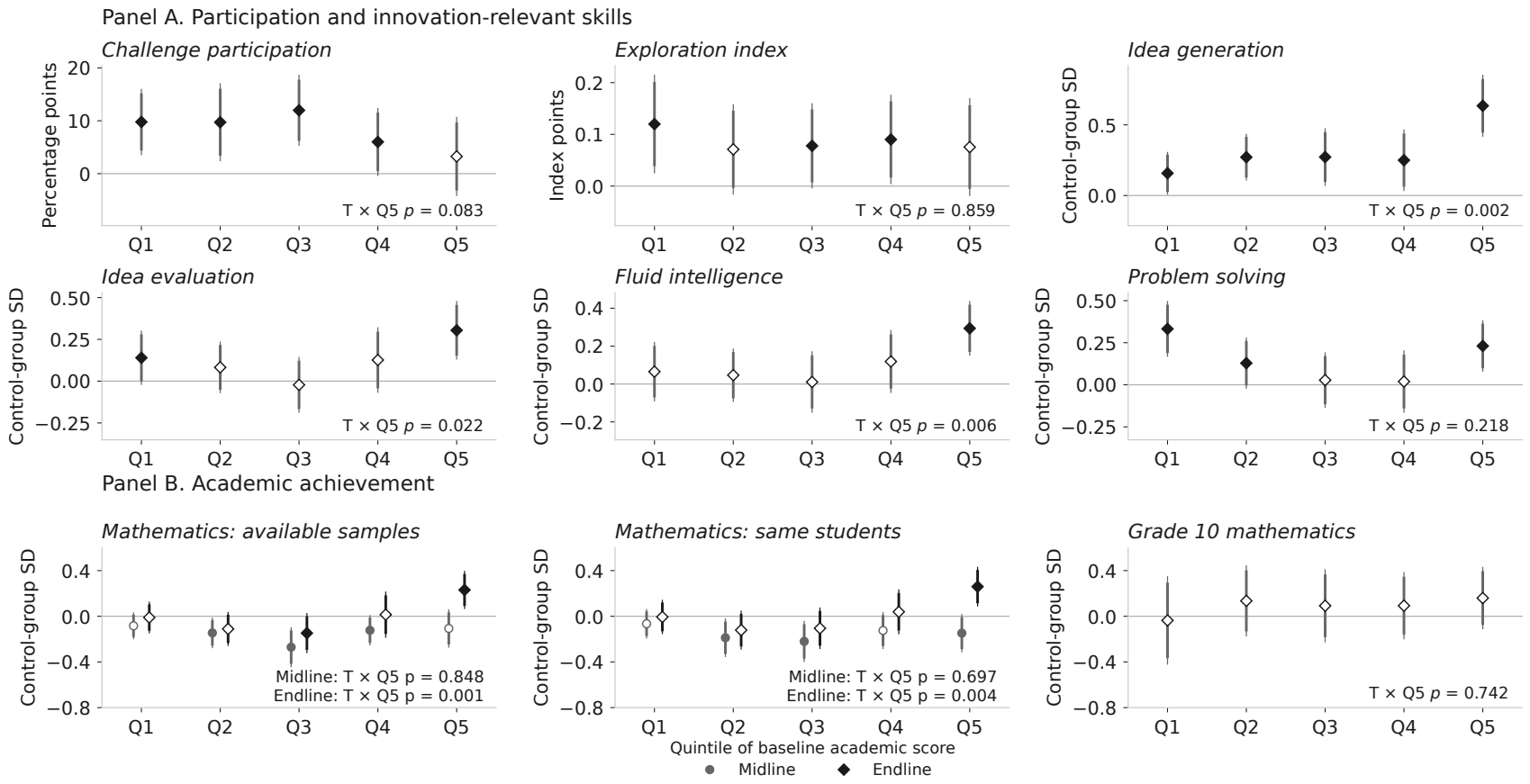


FIGURE 5. TREATMENT EFFECTS BY BASELINE ACADEMIC SCORE QUINTILE

Notes: Points show within-quintile treatment effects; thick and thin bars show 90 and 95 percent confidence intervals based on school-clustered standard errors. Filled markers indicate $p < 0.10$; the gray line marks zero. Panel A reports participation and innovation-relevant skills for 4,388 endline-assessed students. Panel B reports academic achievement. Its left plot compares mathematics effects using the available samples at midline ($N = 4{,}620$) and endline ($N = 4{,}388$); its center plot uses the same 3,641 students at both waves; its right plot reports grade 10 mathematics effects for 5,632 students. Circles identify midline estimates and diamonds identify endline estimates in the two wave comparisons. Quintiles use the combined score from the 27 baseline mathematics and science items, formed within each estimation sample and held fixed across waves in the same-student comparison. Missing baseline scores (22.8 percent at midline, 25.3 percent at endline, 20.8 percent in the same-student sample, and 28.6 percent in the grade 10 sample) are imputed using the mean score within the student's school section, with missing-score indicators included in the regressions. Participation effects are in percentage points, exploration effects in index points, and the remaining effects in control-group standard deviations. Mathematics outcomes retain their original wave-specific control-group standard deviations in the same-student comparison. Regressions include randomization strata, school infrastructure, individual covariates, and the outcome-specific baseline controls and missing-value indicators used in Tables 4–6. The reported interaction p-values compare Q5 with Q1–Q4 pooled within each sample and assessment wave; they do not test differences between waves.

pattern detected.[25]

Taken together, the curriculum expanded participation and developed skills among students with lower academic achievement, while academically stronger students made larger gains on some tasks. The mathematics results nevertheless point to possible academic costs at some stages and for some students. Appendix Tables A.7 and A.8 report the subgroup estimates for every outcome in Tables 4–6, Appendix Figure A.5 shows the outcomes not included in Figure 4, and Appendix Table A.9 reports the joint-interaction and within-subsample estimates.

Direction of Ideas. The problems that students choose to work on differ by their social background, in both arms and in the pooled sample of 1,225 scored submissions (Appendix Figure A.6). Submissions from girls' schools are concentrated in public welfare, health, and households, which account for 32.9 percent of their submissions against 23.6 percent from boys' schools, and submissions from boys' schools are concentrated in the environment and transport. Submissions from tribal-welfare schools are concentrated in agriculture and transport, and submissions from social-welfare schools in the environment and household life. Both patterns appear within the control arm and within the treatment arm, and the pooled differences are statistically significant and remain so after a sharpened false-discovery-rate adjustment across the figure's tests. Teams above and below the median baseline academic score do not differ significantly in the sectors or intended users of their submissions.

Submissions from treatment schools show a greater focus on school-related problems. The curriculum's problem-identification module instructs students to observe their surroundings, interview prospective users, and verify that a problem exists; for students in residential schools, the school provides an immediate setting for such observation. Submissions from treatment schools are directed to farmers less often than submissions from control schools (8.7 against 14.8 percent) and to teachers and schools more often (4.8 against 1.4 percent).

Robustness. The estimated effects retain their signs across specifications ranging from treatment alone to a saturated stratum-cell fixed-effects model (Appendix Tables A.10–A.12, covering the outcomes in Tables 4–6). The idea-rating effects are similar in magnitude with and without controls, and replacing the three strata indicators with the eight stratum cells leaves every coefficient essentially unchanged. The fluid-intelligence estimate remains close to 0.11 standard deviations across specifications, while the problem-solving estimate ranges from 0.10 to 0.14. Covariate adjustment improves precision for both outcomes.

[25] Grade 10 scores are compressed near the top of the scale (Appendix Figure E.1), limiting the scope to detect gains among higher-scoring students. The absence of a detectable subgroup pattern therefore does not establish that the endline differences disappeared.

VII. Cost Effectiveness and Implications for Scale

The estimated financial cost of the two-year program was approximately $80 (₹6,600) per student, or about $40 per student per year, at the 40-school scale evaluated. The estimate covers classroom materials, student-leader training, delivery, monitoring, program management, and an allocation for curriculum development. Per treatment school, the program cost approximately $2,800 (₹2.3 lakh) per year.

Across the four population-level measures of idea quality in Table 4, treatment effects of 0.109 to 0.187 standard deviations imply a cost of approximately $43 to $73 per 0.1 standard deviation of improvement in rated idea quality. A relevant fiscal benchmark is expenditure on India's school innovation infrastructure. The Atal Tinkering Labs program provides selected schools with approximately $24,000 (₹20 lakh) for laboratory establishment and operation over five years, about $4,800 per school per year (Atal Innovation Mission 2024). The programs differ in inputs and student coverage, so this comparison indicates budgetary scale rather than relative cost-effectiveness.

VIII. Discussion and Conclusion

This paper shows that schools can develop students' skills in generating and developing new ideas. In a randomized evaluation across 80 government residential schools in Telangana, India, a two-year curriculum increased participation in an innovation competition, improved the rated quality of student ideas, and developed innovation-relevant skills. Although the instructional sequence and materials were standardized, students chose the problems they addressed and the solutions they developed. The findings establish that structured instruction can improve open-ended production, providing one component of the exposure to innovation that families and neighborhoods supply unevenly (Bell et al. 2019).

The findings have five implications for education research and policy.

First, the curriculum developed individual skills beyond the ideas students produced for the end-of-year Innovation Challenge. Idea generation, exploration, fluid intelligence, and well-defined problem solving were each measured individually, on established tasks unrelated to the teams' projects, and the gains extend across the baseline academic distribution. These skills are therefore malleable at this age in a low-income government school system, and individual students, not only their teams, improved.

Second, innovative production can be measured at the school stage, before patents or commercialization exist. The study scored every team's idea on its novelty and its value to intended users, with a rubric developed with practicing innovators and applied by evaluators blind to treatment status. Two checks support the measure: the treatment effect holds under each evaluator's scores taken separately, and the professional jurors who scored the finalists on the same rubric allocated actual funding in line with their scores.

A complementary measurement contribution is a scalable framework for assessing early-stage innovation. The principal constraint on measuring idea generation at scale is the cost of grading. Open-ended creativity tasks, such as the Alternate Uses Task (Guilford 1967), are simple to administer but require extensive human grading, and many research studies and large-scale assessments therefore avoid them and other open-text measures, even in core subjects such as literacy and numeracy. The scoring validation reported in Section IV.C indicates that AI-based grading of open-ended responses can provide reliable evidence in large samples, where scores are analyzed in aggregate and carry no consequences for individual students, and contributes to the early evidence on such grading.

Both sets of measures are also being adopted beyond this study. With support from the Agency Fund, both the measure of student ideas and the standardized assessment battery are being extended over the coming years to participants in the national School Innovation Marathon, whose 2025 round drew 327,599 submissions (Inqui-Lab Foundation and Gupta forthcoming).

Third, talent identification for innovation depends on preparation, not only on access. Governments identify such talent through competitions and examinations, including innovation challenges, mathematics and science Olympiads, and entrance examinations. These instruments are effective: Olympiad medalists win major science prizes at ten times the rate of undergraduates at the world's leading universities (Agarwal, Gaulé, and Jiang 2025). They identify talent, however, only among entrants, and access alone does not produce entry. In this experiment, every school was offered the same competition, and approximately one quarter of control students did not participate. Instruction raised participation by 8 percentage points, and the additional participants were disproportionately students with lower academic achievement. The curriculum also raised the rated novelty and usefulness of the ideas produced. In the absence of preparation, a competition therefore identifies talent among the students who would enter without preparation, and the students it excludes are disproportionately those whom examination-based identification also excludes. Preparation in school extends identification to these students.

Fourth, academic performance provides an incomplete basis for identifying innovation-relevant talent. In many school systems, examination scores determine entry into science tracks and selective technical programs. In India, grade 10 examination scores influence entry into science streams (e.g., TGSWREIS 2025, p. 2), and entrance examinations govern admission to technical degrees. The incentives surrounding these examinations can direct instruction toward examination performance rather than skills with broader applications (Muralidharan 2024). Among control students, baseline academic scores explained at most 12 percent of variation in the assessed innovation-relevant skills, and grade 10 examination scores at most 10 percent. Academic rankings therefore provide an incomplete account of the skills measured here. Developing

these skills has a role both before academic selection and among students who advance through it.

Fifth, innovation instruction can be delivered without specialist classroom teachers or dedicated laboratories. The delivery model offers a concrete approach to advancing the creativity and innovation objectives of India's National Education Policy (Government of India 2020). Two student leaders per classroom received 20 hours of training and led sessions using structured workbooks, with teacher support and regular external monitoring. Specialist expertise entered through curriculum design, training, and continuing support, without requiring specialist classroom teachers or dedicated laboratories. At approximately $40 per student annually, the program produced these gains without a detectable average reduction in endline academic achievement or subsequent state graduation examination scores. Academic costs at some stages and for some students nevertheless cannot be ruled out.

These findings suggest a role for schools in innovation policy, alongside investments in research infrastructure, finance, and entrepreneurship support. Preparing students to formulate and test ideas may improve their use of these later opportunities and eventually expand participation in the search for new ideas. Broader participation may also affect which problems receive attention. Girls' and boys' teams, and teams from tribal- and social-welfare schools, proposed ideas in different sectors. These descriptive differences are consistent with evidence linking inventors' backgrounds to the direction of innovation (Bell et al. 2019; Koning, Samila, and Ferguson 2021; Einiö, Feng, and Jaravel 2026).

Further work could examine how improvements in individual skills contribute to team production and whether the gains can be sustained with less external monitoring or with teacher delivery integrated into the program. Following students beyond school could test whether these gains persist and affect subsequent innovation or entrepreneurship. The present findings establish that schooling can develop human capital for innovation, with gains extending beyond the students whom academic performance would identify as most promising.

REFERENCES

Agarwal, Ruchir, and Patrick Gaulé. 2020. "Invisible Geniuses: Could the Knowledge Frontier Advance Faster?" *American Economic Review: Insights* 2 (4): 409–424.

Agarwal, Ruchir, Patrick Gaulé, and Yuyan Jiang. 2025. "Finding Young Einsteins: Olympiads and STEM Talent Discovery." Global Talent Fund Working Paper WP/25/01.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth through Creative Destruction." *Econometrica* 60 (2): 323–351.

Aghion, Philippe, Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen. 2017. "The Social Origins of Inventors." National Bureau of Economic Research Working Paper 24110, Cambridge, MA.

Aghion, Philippe, Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen. 2023. "2022 Klein Lecture Parental Education and Invention: The Finnish Enigma." *International Economic Review* 64 (2): 453–490.

Alan, Sule, and Elif Kubilay. 2025. "Empowering Adolescents to Transform Schools: Lessons from a Behavioral Targeting." *American Economic Review* 115 (2): 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): 1173–1210.

Alan, Sule, Teodora Boneva, and Seda Ertac. 2019. "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit." *Quarterly Journal of Economics* 134 (3): 1121–1162.

Al-Ubaydli, Omar, John A. List, and Dana L. Suskind. 2017. "What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results." *American Economic Review* 107 (5): 282–286.

Amaral, Sofia, Aakash Bhalothia, Ritam Chaurey, Isis Gaddis, Gaurav Khanna, Samreen Malik, Abhiroop Mukhopadhyay, Nishith Prakash, and Raghav Rakesh. 2025. "Building Entrepreneurial Mind-Sets: Evidence from a Large-Scale Educational Intervention in India." World Bank Group South Asia Gender Innovation Lab Policy Brief, Washington, DC. https://documents.worldbank.org/curated/en/099520405072521933.

Anderson, Lorin W., and David R. Krathwohl, eds. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives.* New York: Longman.

ASER Centre. 2025. "Annual Status of Education Report (Rural) 2024." ASER Centre, New Delhi. Provisional report, released January 28, 2025. https://asercentre.org/aser-2024/ (accessed September 26, 2026).

Asher, Sam, Paul Novosad, and Charlie Rafkin. 2024. "Intergenerational Mobility in India: New Measures and Estimates Across Time and Social Groups." *American Economic Journal: Applied Economics* 16 (2): 66–98.

Ashraf, Nava, Natalie Bau, Corinne Low, and Kathleen McGinn. 2020. "Negotiating a Better Future: How Interpersonal Skills Facilitate Intergenerational Investment." *Quarterly Journal of Economics* 135 (2): 1095–1151.

Ashraf, Nava, Natalie Bau, Corinne Low, and Xiaoyue Shan. 2025. "Ten Years of Relational Power: The Long-Run Effects of Teaching Negotiation Skills to Adolescent Girls." NBER Working Paper 34339. Forthcoming, *AER: Insights*.

Atal Innovation Mission. 2024. "ATL Guidebook: Integrating Innovation in School Education." Atal Innovation Mission, NITI Aayog. https://aim.gov.in/pdf/ATL-Guidebook.pdf (accessed September 26, 2026).

Atal Innovation Mission. 2026. "Atal Tinkering Labs." Program webpage. https://aim.gov.in/atl.php (accessed October 9, 2026).

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 26280.

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): 241–251.

Bell, Alexander, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen. 2019. "Who Becomes an Inventor in America? The Importance of Exposure to Innovation." *Quarterly Journal of Economics* 134 (2): 647–713.

Bharti, Nitin Kumar. 2018. "Wealth Inequality, Class and Caste in India, 1961–2012." World Inequality Lab Working Paper 2018/14.

Blimpo, Moussa P., and Todd Pugatch. 2021. "Entrepreneurship Education and Teacher Training in Rwanda." *Journal of Development Economics* 149: 102583.

Bloom, Benjamin S., Max D. Engelhart, Edward J. Furst, Walker H. Hill, and David R. Krathwohl. 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain.* New York: David McKay.

Bloom, Nicholas, John Van Reenen, and Heidi Williams. 2019. "A Toolkit of Policies to Promote Innovation." *Journal of Economic Perspectives* 33 (3): 163–184.

Brown, Christina, Supreet Kaur, Geeta Kingdon, and Heather Schofield. 2025. "Cognitive Endurance as Human Capital." *Quarterly Journal of Economics* 140 (2): 943–1002.

CGD. 2025. "The Big Returns to Early Investments in Foundational Literacy and Numeracy." Center for Global Development event and transcript, October 17. https://www.cgdev.org/event/big-returns-early-investments-foundational-literacy-and-numeracy.

Chioda, Laura, Paul J. Gertler, David Contreras-Loya, and Dana Rose Carney. 2026. "Making Entrepreneurs: Long Term Returns to Training Youth in Business Skills." NBER Working Paper 34637.

Commonwealth of Kentucky, Office of State Budget Director. 2024. "2024–2026 Budget of the Commonwealth." Volume I, p. 103. https://osbd.ky.gov/Publications/Documents/Budget%20Documents/2024-2026%20Budget%20of%20the%20Commonwealth/2024-2026%20Budget%20of%20the%20Commonwealth%20-%20Volume%20I.pdf (accessed October 12, 2026).

Condon, David M., and William Revelle. 2014. "The International Cognitive Ability Resource: Development and

Initial Validation of a Public-Domain Measure." *Intelligence* 43: 52–64.

Connecticut Department of Economic and Community Development. 2026. "Manufacturing Innovation Fund: FY2025 Annual Report." pp. 21–24. https://manufacturing.ct.gov/-/media/manufacturing/fy25-mif-annual-report-accessible.pdf (accessed October 9, 2026).

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47.

Dejuan, Irene, and Elisa Sanz Navarro. 2023. "Process Evaluation of the School Innovation Challenge Programme in Telangana, India." UNICEF Regional Office for South Asia. Prepared by the Center for Evaluation and Development (C4ED). https://evaluationreports.unicef.org/GetDocument?documentID=19421&fileID=51439 (accessed September 26, 2026).

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market." *Quarterly Journal of Economics* 132 (4): 1593–1640.

Deming, David J. 2022. "Four Facts about Human Capital." *Journal of Economic Perspectives* 36 (3): 75–102.

EAST Initiative. 2024. "Application period now open for EAST Initiative grant funding." August 22. https://www.eastinitiative.org/images/slider/2024-25_GrantApplication.pdf (accessed October 12, 2026).

Ederer, Florian, and Gustavo Manso. 2013. "Is Pay for Performance Detrimental to Innovation?" *Management Science* 59 (7): 1496–1513.

Einiö, Elias, Josh Feng, and Xavier Jaravel. 2026. "Social Push and the Direction of Innovation." *American Economic Review* 116 (10): 3875–3909.

Evans, David, and Susannah Hares. 2021. "Should Governments and Donors Prioritize Investments in Foundational Literacy and Numeracy?" Center for Global Development. May 4. https://www.cgdev.org/publication/should-governments-and-donors-prioritize-investments-foundational-literacy-and-numeracy.

Gneezy, Uri, and Jan Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *Quarterly Journal of Economics* 112 (2): 631–645.

Government of India. 2020. "National Education Policy 2020." Ministry of Human Resource Development, New Delhi. https://static.pib.gov.in/WriteReadData/userfiles/NEP_Final_English_0.pdf.

Government of Telangana. 2024. "Deputy CM presents Budget for the year 2024–25." Telangana State Portal, July 25. https://www.telangana.gov.in/news/press-releases/2024/07/deputy-cm-presents-budget-for-the-year-2024-25/ (accessed October 12, 2026).

Government of Telangana. 2026. "Telangana Gurukulam Common Entrance Test 2026: Results." https://tgcet.cgg.gov.in/ (accessed July 6, 2026).

Guilford, J. P. 1967. *The Nature of Human Intelligence.* New York: McGraw-Hill.

Gupta, Saloni. 2025. "Can Innovation Be Taught in Schools? Experimental Evidence from India." *AEA RCT Registry*, May 26. https://doi.org/10.1257/rct.9696.

Heckman, James J., and Tim Kautz. 2012. "Hard Evidence on Soft Skills." *Labour Economics* 19 (4): 451–464.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24 (3): 411–482.

Hoisl, Karin, Hans Christian Kongsted, and Myriam Mariani. 2023. "Lost Marie Curies: Parental Impact on the Probability of Becoming an Inventor." *Management Science* 69 (3): 1714–1738.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow. 2019. "The allocation of talent and US economic growth." *Econometrica* 87 (5): 1439–1474.

Inqui-Lab Foundation, and Saloni Gupta. forthcoming. "School Innovation Marathon 2025 Report." Manuscript.

Iowa Department of Education. 2026. "STEM BEST Program." Program webpage, 2026 funding cycle. https://educate.iowa.gov/iowa-stem/stem-best (accessed October 9, 2026).

Jha, Anushka. 2025. "As Shark Tank India Turns Five, Sony Bets on Scale across TV, OTT and Connected Screens." *afaqs!*, December 19. https://www.afaqs.com/marketing/as-shark-tank-india-turns-five-sony-bets-on-scale-across-tv-ott-and-connected-screens-10922444 (accessed July 22, 2026).

Jones, Charles I. 2022. "The Past and Future of Economic Growth: A Semi-Endogenous Perspective." *Annual Review of Economics* 14: 125–152.

Kenya Institute of Curriculum Development. 2024. "Junior School Curriculum Design: Pre-Technical Studies, Grade 9." Kenya Institute of Curriculum Development. ISBN 978-9914-43-985-4. https://kicd.ac.ke/cbc-materials/curriculum-designs/grade-nine-designs/ (accessed September 27, 2026).

Khan, Saniya Ahmad. 2024. "Shark Tank India: Impact on India's Growing Startup Mindset." *YourStory*, February 9. https://yourstory.com/2024/02/shark-tank-india-startup-impact (accessed September 26, 2026).

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): 1345–1348.

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

List, John A. 2024. "Optimally Generate Policy-Based Evidence Before Scaling." *Nature* 626 (7999): 491–499.

Manso, Gustavo. 2011. "Motivating Innovation." *Journal of Finance* 66 (5): 1823–1860.

March, James G. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science* 2 (1): 71–87.

Massachusetts Life Sciences Center. 2025. "Healey-Driscoll Administration Announces $12.7 Million to Support Life Sciences Education and Workforce Development." October 17. https://www.masslifesciences.com/healey-driscoll-administration-announces-12-7-million-to-support-life-sciences-education-and-workforce-development/ (accessed October 9, 2026).

Ministry of Education, Government of India. 2025. "Increasing Dropout Rates." Annexure to Rajya Sabha Unstarred Question No. 348, answered July 23, pp. 1–3. https://www.education.gov.in/sites/upload_files/mhrd/files/parliament_annexure_en/rsuq_348.pdf (accessed October 9, 2026).

Munshi, Kaivan. 2019. "Caste and the Indian Economy." *Journal of Economic Literature* 57 (4): 781–834.

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

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

Muralidharan, Karthik, and Abhijeet Singh. 2025. "Adapting for Scale: Experimental Evidence on Technology-Aided Instruction in India." NBER Working Paper 34205.

Muralidharan, Karthik, Jishnu Das, Alaka Holla, and Aakash Mohpal. 2017. "The Fiscal Cost of Weak Governance: Evidence from Teacher Absence in India." *Journal of Public Economics* 145: 116–135.

Newell, Allen, and Herbert A. Simon. 1972. *Human Problem Solving.* Englewood Cliffs, NJ: Prentice-Hall.

Nourani, Vesall, Nava Ashraf, and Abhijit Banerjee. 2025. "Learning to Teach by Learning to Learn." Working paper.

OECD. 2024. *PISA 2022 Results (Volume III): Creative Minds, Creative Schools.* Paris: OECD Publishing. https://doi.org/10.1787/765ee8c2-en.

OECD and Eurostat. 2018. *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation. The Measurement of Scientific, Technological and Innovation Activities*. 4th ed., Paris: OECD Publishing. https://doi.org/10.1787/9789264304604-en.

Press Information Bureau. 2024. "School Innovation Marathon 2024 Attracts Over 1 Lakh Submissions." Press release, NITI Aayog, December 20. https://www.pib.gov.in/PressReleseDetailm.aspx?PRID=2086453 (accessed July 4, 2026).

Press Information Bureau. 2025. "Shri Dharmendra Pradhan Launches Viksit Bharat Buildathon 2025, to Ignite Innovation among School Students across the Country." Press release, Ministry of Education, September 23. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2170192 (accessed August 27, 2026).

Pritchett, Lant, and Amanda Beatty. 2015. "Slow Down, You're Going Too Fast: Matching Curricula to Student Skill Levels." *International Journal of Educational Development* 40: 276–288.

Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5, Part 2): S71–S102.

SCERT Delhi. 2021. "Annual Report (2019–2020)." State Council of Educational Research and Training, Government of NCT of Delhi, New Delhi.

Sihag, Garima. 2026. "TGCET Gurukul Result 2026 Out at tgcet.cgg.gov.in: Check 5th, 9th Score Card, Rank." *Careers360*, updated May 15. https://school.careers360.com/articles/tg-gurukul-tgcet-result-2026 (accessed September 26, 2026).

Singh, Abhijeet. 2024. "Improving Administrative Data at Scale: Experimental Evidence on Digital Testing in Indian Schools." *Economic Journal* 134 (661): 2207–2223.

SonyLIV. 2026. "Shark Tank India." https://www.sonyliv.com/shows/shark-tank-india-1700000741 (accessed June 30, 2026).

Telangana State Innovation Cell. 2026. "UNICEF-TSIC School Innovation Challenge." https://teamtsic.telangana.gov.in/unicef-tsic-school-innovation-challenge/ (accessed July 4, 2026).

TGSWREIS. 2025. "Admission into Junior Intermediate in COE, Sainik and Non COE Colleges (General & Vocational) for the Academic Year 2025–26." Telangana Social Welfare Residential Educational Institutions Society. Rc.No.Acad-I/25/TGCET/2025, April 24. https://drive.google.com/file/d/1AMSY8qvoefPgde_zNlL94ctZxvm2T

M8X/view (accessed October 16, 2026).

TGSWREIS. 2026. "About Us." Telangana Social Welfare Residential Educational Institutions Society. https://tgswreis.telangana.gov.in/ (accessed September 26, 2026).

TGTWREIS. 2024. "Academic Results from the Year 2013–2024." Telangana Tribal Welfare Residential Educational Institutions Society. https://tgtwreis.telangana.gov.in/content/results-670e9754c0ecb (accessed July 11, 2026).

TGTWREIS. 2026. "Telangana Tribal Welfare Residential Educational Institutions Society." Official website. https://tgtwreis.telangana.gov.in/ (accessed September 26, 2026).

Toivanen, Otto, and Lotta Väänänen. 2016. "Education and Invention." *Review of Economics and Statistics* 98 (2): 382–396.

TSWREIS. 2022. "TSWREIS Students Excel in SSC Exam." Telangana Social Welfare Residential Educational Institutions Society. https://tgswreis.telangana.gov.in/slider1-SSC-2022.php (accessed July 5, 2026).

Van Reenen, John. 2021. "Innovation and Human Capital Policy." NBER Working Paper 28713.

Weidmann, Ben, and David J. Deming. 2021. "Team Players: How Social Skills Improve Team Performance." *Econometrica* 89 (6): 2637–2657.

Wisconsin Economic Development Corporation. 2026. "Fabrication Laboratories (Fab Lab)." Program summary. https://wedc.org/fabrication-laboratories/ (accessed October 9, 2026).

World Bank. 2024a. "FLEX 2024: African Ministers and Education Stakeholders Gather in Kigali to Track Progress in Addressing Learning Crisis on the Continent." Press release, November 11. https://www.worldbank.org/en/news/press-release/2024/11/11/flex-2024-afe-ministers-and-education-stakeholders-gather-in-kigali-to-track-progress-in-addressing-learning-crisis.

World Bank. 2024b. *World Development Report 2024: The Middle-Income Trap.* Washington, DC: World Bank.

World Bank. 2026a. "Research and Development Expenditure (% of GDP)." *World Development Indicators*, Washington, DC: World Bank. https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS (accessed August 27, 2026).

World Bank. 2026b. "Researchers in R&D (per Million People), 2020." *World Development Indicators*, Washington, DC: World Bank. https://data.worldbank.org/indicator/SP.POP.SCIE.RD.P6?locations=IN (accessed August 4, 2026).

World Intellectual Property Organization. 2024. "World Intellectual Property Report 2024: Making Innovation Policy Work for Development." World Intellectual Property Organization, Geneva. https://doi.org/10.34667/tind.49284.

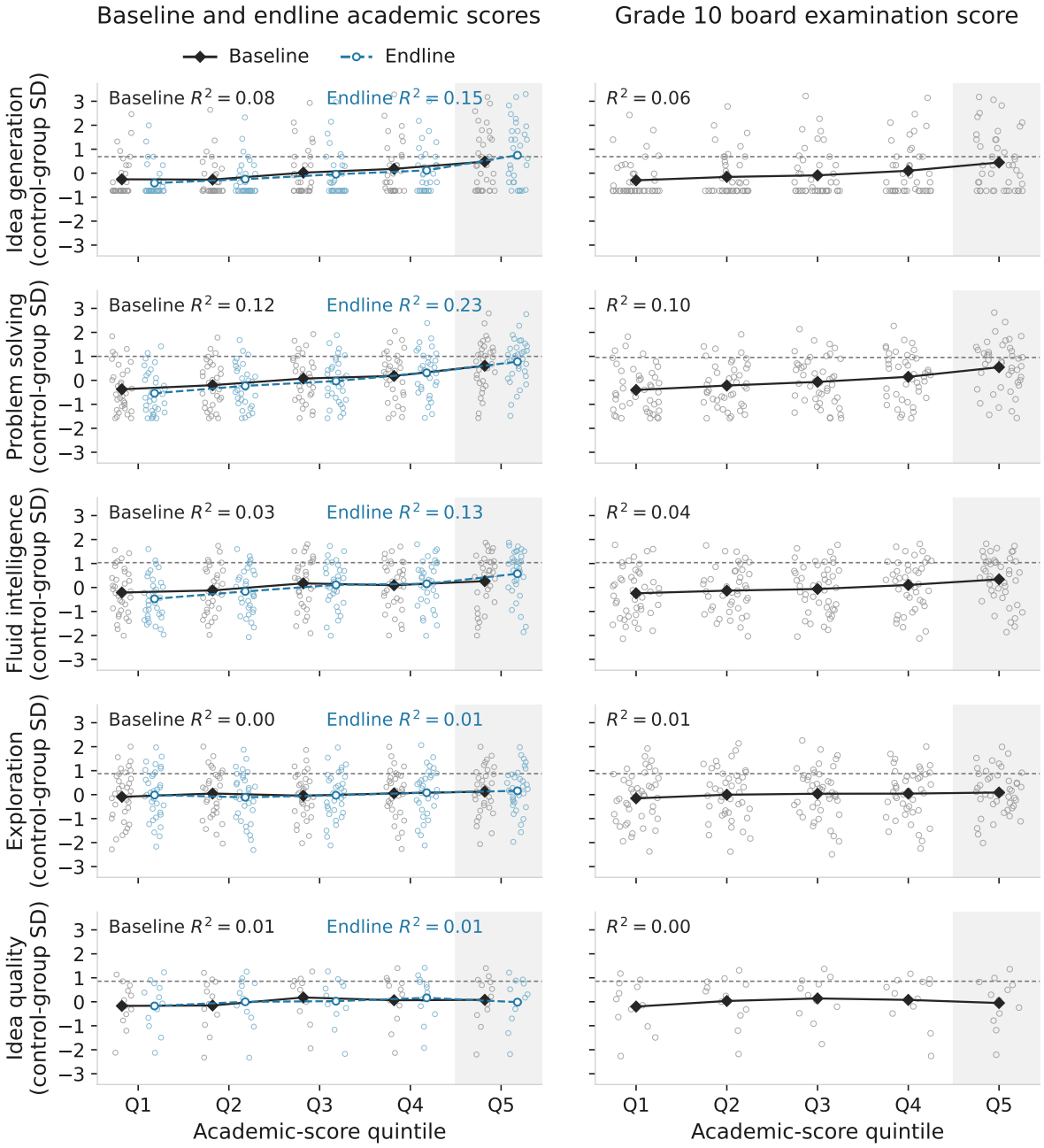


FIGURE 1. ACADEMIC SCORES AND INNOVATION-RELEVANT SKILLS

Notes: The first four rows plot endline skill measures among control students by academic-score quintile. The left column compares baseline and endline rankings for the same 1,602 students with observed scores; the right uses grade 10 board totals ($N = 2{,}123$). Baseline scores use pooled mathematics–science IRT; endline scores average standardized mathematics and science scores. The final row plots idea quality (Innovation-S) for 554 scored submitting control teams against team academic scores, averaging equally weighted mathematics–science composites over members with observed scores at each wave. Imputed baseline scores are excluded, and team quintiles are defined across teams. Quintiles are defined separately for each assessment. Each outcome is standardized within the relevant control group. Gray and blue points summarize approximately ten adjacent outcome-ranked students or teams within each quintile; black diamonds and blue circles mark baseline and endline quintile means, respectively. In the right column, gray points and black diamonds summarize the board-score rankings. The shaded region marks the highest academic quintile; the horizontal dashed line marks the outcome's 80th percentile, including ties. Panel R^2 values use continuous academic scores.

Appendix

A. Additional Tables and Figures

TABLE A.1—SAMPLE AND INNOVATION CHALLENGE FUNNEL

	Control	Treatment	Total
Panel A. Individual sample, by wave			
Study cohort (grade 10 examination rosters)	2,807	2,835	5,642
Baseline assessment	2,597	2,618	5,215
Midline assessment	2,301	2,319	4,620
Endline assessment	2,126	2,262	4,388
Grade 10 board records	2,801	2,832	5,633
Panel B. Endline Innovation Challenge (team analysis)			
Students			
Assessed at endline	2,126	2,262	4,388
On a submitted team	1,558	1,828	3,386
Non-submitters (not on a submitted team)	568	434	1,002
Teams			
Non-submitter teams (grouped ≤ 4 within school $\times$ arm)	155	122	277
Idea-rating estimation sample	723	770	1,493

Notes: Panel A reports the number of students in each data source, by treatment arm. The first row counts the students on the grade 10 board examination rosters of the 80 study schools. The baseline, midline, and endline rows count the students assessed in each wave. The grade 10 row counts the students with a board examination result (pass or fail). Panel B reports the construction of the Innovation Challenge team sample. Students assessed at endline are divided into those on a submitted team and non-submitters. Non-submitters form constructed teams of up to four students within school; estimation follows Section V. The idea-rating estimation sample of 1,493 teams combines the 1,216 scored submitted teams that have at least one member assessed at endline with the 277 constructed non-submitter teams.

TABLE A.2—SCHOOL CHARACTERISTICS AND BALANCE

	Control mean	Treatment mean	Treatment− Control	*p*-value	*N*
Panel A. Randomization stratification characteristics					
Girls-only school	0.500	0.500	0.000		80
Tribal-welfare school	0.500	0.500	0.000		80
Rural school	0.625	0.625	0.000		80
Panel B. School infrastructure					
Rented building	0.250	0.375	0.125	0.233	80
Has library	0.500	0.500	0.000	1.000	80
Has playground	0.125	0.150	0.025	0.749	80
English medium	0.825	0.850	0.025	0.765	80
Has ramp	0.725	0.675	-0.050	0.631	80
Has computer	0.675	0.650	-0.025	0.816	80
Boys' toilets	6.175	6.800	0.625	0.806	80
Girls' toilets	9.250	8.300	-0.950	0.758	80
Degraded boundary wall	0.075	0.075	0.000	1.000	80
No boundary wall	0.175	0.175	0.000	1.000	80

Notes: The unit of observation is the school ($N = 80$). School characteristics are drawn from the Unified District Information System for Education (UDISE), the administrative school records used for randomization, and were recorded before the baseline assessment in 2022. Panel A reports the three characteristics used to stratify random assignment. The complementary categories are boys-only schools, social-welfare schools, and urban schools, respectively. The sample contains 40 girls-only and 40 boys-only schools, and 40 social-welfare and 40 tribal-welfare schools; each treatment arm contains 20 schools from each of these categories. Panel B reports pre-treatment school infrastructure. Each row reports means by treatment arm and the treatment-control difference. Panel A is descriptive, so *p*-values are omitted. In Panel B, differences are estimated from school-level regressions of each characteristic on treatment assignment, with heteroskedasticity-robust standard errors. The two toilet rows report the number of toilets; all other characteristics are binary indicators.

TABLE A.3—SAMPLE COVERAGE AND ATTRITION

	Control mean	Treatment mean	Treatment − Control	*p*-value	*N*
Panel A. Observation rates among baseline-assessed students (%)					
Observed in midline assessment	69.6	67.2	-2.2	0.231	5,215
Observed in endline assessment	61.7	63.9	2.4	0.292	5,215
Linked to grade 10 records	77.7	76.8	-0.7	0.677	5,215
Panel B. Differential attrition: treatment × baseline covariate on endline observation					
Treatment × Age			2.06	0.122	5,215
Treatment × Mathematics			-0.30	0.850	5,215
Treatment × Science			-1.44	0.397	5,215
Treatment × Fluid intelligence			0.40	0.791	5,215
Treatment × Emotional intelligence			0.11	0.951	5,215
Treatment × Big Five			1.82	0.197	5,215
Treatment × Risk preference			0.61	0.644	5,215
Joint test *p*-value				0.501	5,215
Attriter difference: Grade 10 mathematics (ext., SD)			0.036	0.580	5,632
Attriter difference: Grade 10 science (ext., SD)			-0.026	0.572	5,632
Panel C. Baseline balance among endline-observed students					
Age (years)	13.261	13.246	-0.037	0.421	3,276
Household SES index (SD)	0.061	-0.056	-0.130	0.156	2,884
Household computer literacy (0–1)	0.378	0.418	0.048	0.109	2,884
Mathematics score (%)	42.8	42.2	-0.1	0.953	3,276
Science score (%)	53.4	53.3	0.3	0.801	3,276
Fluid intelligence (% correct)	59.5	59.8	0.3	0.829	3,276
Emotional intelligence (% correct)	53.4	52.8	-0.3	0.737	3,276
Big Five personality traits (mean item response)					
Agreeableness	3.251	3.260	0.015	0.544	3,276
Conscientiousness	3.254	3.232	-0.015	0.498	3,276
Extraversion	3.162	3.165	0.009	0.627	3,276
Neuroticism	2.882	2.878	-0.004	0.806	3,276
Openness	3.228	3.225	0.008	0.788	3,276
Risk preference (% of endowment)	62.2	63.0	0.6	0.534	3,272
Joint test *p*-value				0.804	3,276

Notes: The sample in Panels A and B is the 5,215 students who completed the baseline assessment. Panel A reports the percentage of these students observed in each follow-up data source, by treatment arm. The treatment-control difference is the coefficient on treatment from a regression of the observation indicator on treatment and randomization strata, with standard errors clustered at the school level. Panel B tests whether observation in the endline assessment depends on baseline characteristics differently in the two arms. Each of the first seven rows reports the coefficient on treatment interacted with a standardized baseline covariate, from a single regression of endline observation on treatment, the covariates, their interactions, and randomization strata. Coefficients are in percentage points per standard deviation of the covariate. The joint test is a Wald test that the seven interaction coefficients are zero. Missing covariate values are set to the sample mean in this regression. The last two rows of Panel B report the difference in externally graded grade 10 board examination scores, in standard deviations, between students not observed at endline and students observed, conditional on randomization strata; these rows use all students with a board examination score. Panel C reports baseline covariate means by treatment arm among students observed in the endline assessment, with the treatment-control difference estimated as in Panel A. *N* is the number of students with a nonmissing value in each row. Household SES and computer literacy use observed midline responses, excluding imputed values. The five personality traits in Panel C are mean item responses on a 1–5 scale; neuroticism is reported in its original direction. Variable definitions follow Table 2. Panel C's joint test uses the seven covariates in Panel B, including the Big Five index: the mean of the five trait scores standardized to the control group, with neuroticism reversed.

TABLE A.4—SUBMITTING TEAMS: BASELINE CHARACTERISTICS AND BOUNDS ON ENDLINE IDEA QUALITY

	Control mean	Treatment mean	Treatment−Control	*p*-value	*N*
Panel A. Team-average baseline characteristics					
Members with baseline records (%)	73.7	69.6	-3.4	0.136	1,225
Age (years)	13.294	13.274	-0.042	0.337	1,225
Academic score (SD)	0.012	-0.061	-0.048	0.577	1,225
Fluid intelligence (SD)	0.001	-0.039	-0.034	0.581	1,225
Emotional intelligence (SD)	0.035	-0.051	-0.070	0.285	1,225
Exploration index (SD)	0.014	-0.009	-0.019	0.626	1,225
Risk preference (SD)	-0.029	0.015	0.032	0.461	1,225
Household SES index (SD)	0.044	-0.056	-0.114	0.210	1,225
Household computer literacy (0–1)	0.374	0.394	0.026	0.367	1,225
Joint test *p*-value				0.789	1,225
Team-formation margins					
Team size (members)	3.83	3.90	0.08	0.130	1,225
Within-team SD: academic score	0.684	0.655	-0.025	0.415	1,221
Within-team SD: fluid intelligence	0.774	0.719	-0.051	0.051	1,221
Within-team SD: exploration index	0.483	0.486	0.006	0.771	1,221
Joint test *p*-value				0.061	1,221
Panel B. Lee (2009) bounds on conditional quality differences					
	T−C	Lower	Upper		*N*
Novelty	0.226	0.150	0.380		1,225
Value-Add	0.162	0.082	0.318		1,225
Feasibility	0.080	-0.018	0.218		1,225
Willingness to Fund	0.158	0.047	0.292		1,225

Notes: The sample comprises the 1,225 scored team submissions to the Innovation Challenge. Panel A reports the baseline characteristics of each team, measured as the average across its members, by treatment arm. A missing value for a member is replaced by the mean of that member's school and section. The treatment-control difference is the coefficient on treatment from a regression of the characteristic on treatment and randomization strata, with standard errors clustered at the school level. Each joint test is a Wald test from a regression of treatment on the characteristics in the corresponding block and randomization strata. The within-team standard deviations in the team-formation rows are calculated across members with a baseline record and are defined for teams with at least two such members. Panel B reports Lee (2009) bounds on the treatment-control difference in evaluator ratings among submitted teams. Ratings are expressed in standard deviations of the ratings of submitted control teams, and the differences are not adjusted for randomization strata. In the endline cohort, 84.2 percent of treatment teams and 78.7 percent of control teams submitted an idea. The lower bound therefore excludes the 6.6 percent of treatment teams with the highest ratings, and the upper bound excludes the 6.6 percent with the lowest ratings. The bounds are valid under the assumption that treatment does not prevent any team from submitting. The table reports estimated bound endpoints without confidence intervals.

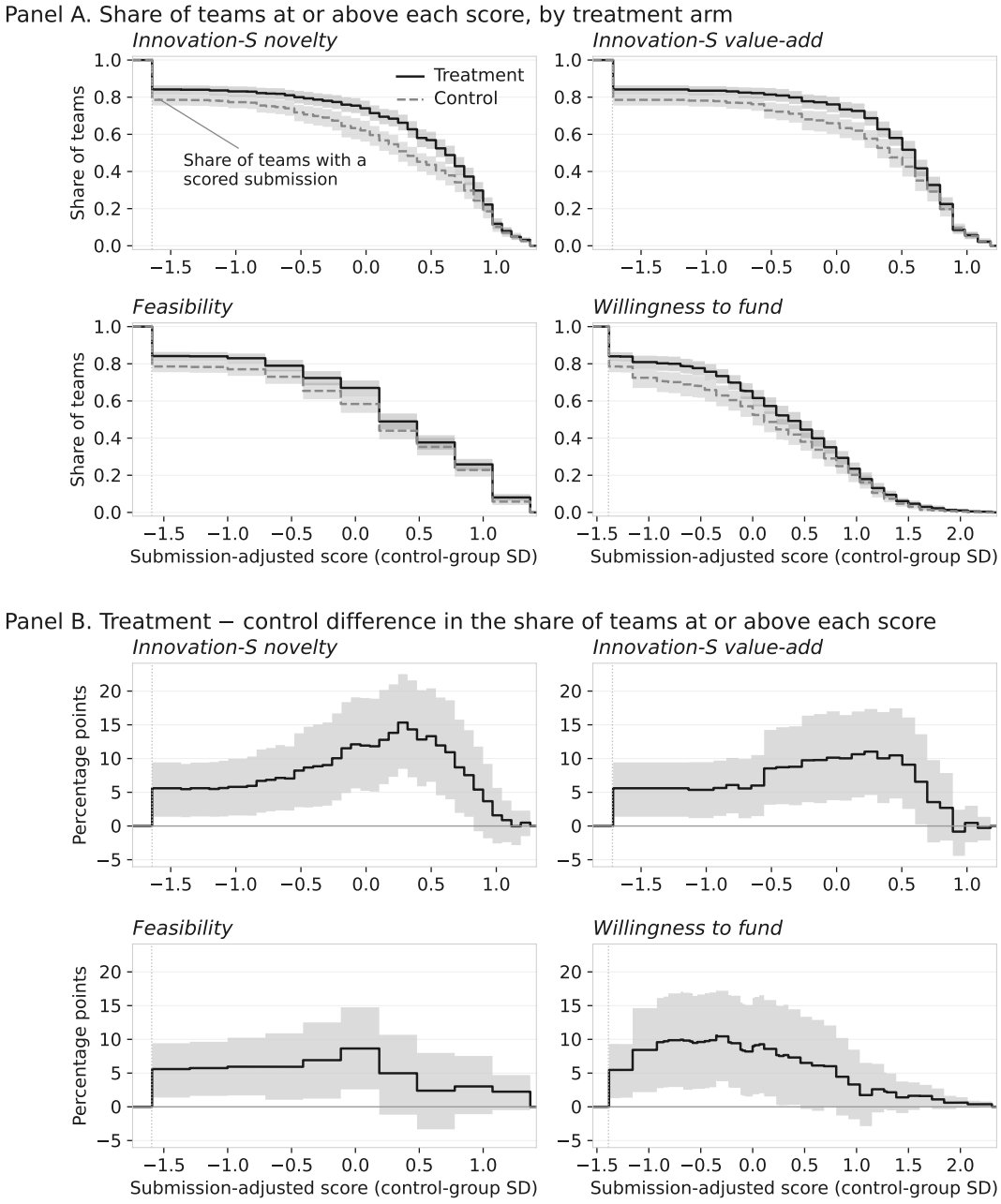


FIGURE A.1. IDEA OUTPUT ACROSS THE DISTRIBUTION OF RATED QUALITY

Notes: The figure reports, for each of the four evaluator ratings, the share of teams whose idea is rated at or above each score. Panel A plots this share separately for treatment and control schools. Panel B plots the treatment-control difference in the share, in percentage points. The sample comprises 1,493 teams: the 1,216 scored submitted teams with at least one member assessed at endline and 277 constructed teams of non-submitters. Teams without a scored submission are assigned the value of no idea, which is marked by the dotted vertical line. Scores are expressed in standard deviations of the control group in this sample. In Panel A, each curve declines at the dotted line by the share of teams without a scored submission. In Panel B, the difference immediately to the right of the dotted line is therefore the difference in the share of teams that submitted an idea, and the difference at higher scores is the difference in the share of teams whose idea meets each rating threshold. Shaded areas are pointwise 95 percent confidence intervals from a bootstrap that resamples schools with replacement within treatment arm, with 999 replications.

TABLE A.5—JURY FUNDING RAISED AT THE SHARK TANK FINALS

Outcome	Control mean	Treatment effect	% of control	*p*-value	*N*
Total jury funding raised, all invited teams	₹6,300	+₹1,709	+27%	0.28	80
Attending teams, with jury-panel fixed effects	₹6,632	+₹1,064	+16%	0.40	78

Notes: The sample comprises the Innovation Challenge finalists, the top-rated teams from each treatment arm invited to the Shark Tank finals. The outcome is the total funding that a team received from its jurors, in rupees. The jurors' allocations were paid to the teams, in contrast to the evaluators' willingness-to-fund rating, which was hypothetical. The first row includes all 80 invited teams. Two control teams were invited but did not attend, and their funding is recorded as zero. The *p*-value in the first row is from randomization inference that reassigns treatment across schools within strata, with 5,000 permutations. It is computed on the 78 teams that attended, because the schools of the two absent teams are not observed. The second row restricts the sample to the 78 attending teams and adds jury-panel fixed effects. Each team faced a single jury panel; these fixed effects account for differences in average generosity across panels. Its *p*-value is based on standard errors clustered at the school level. One jury panel evaluated only control teams, so the second row identifies the effect from the remaining panels. The attending teams are drawn from approximately 29 schools.

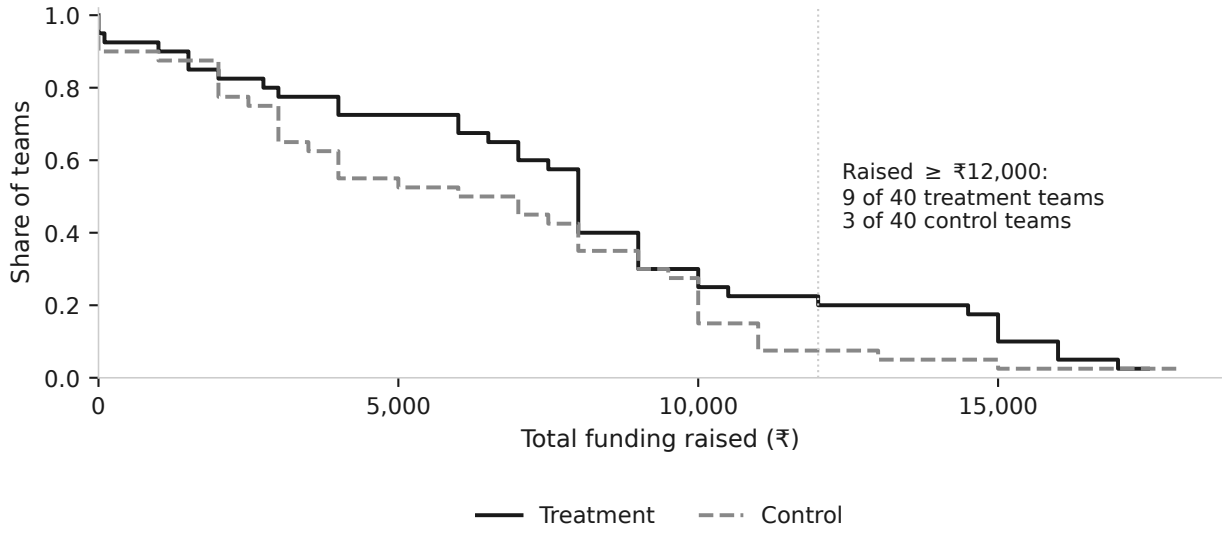


FIGURE A.2. DISTRIBUTION OF JURY FUNDING AT THE SHARK TANK FINALS

Notes: The figure plots, separately by treatment arm, the share of teams that raised at least each funding amount. The sample comprises the 80 invited finalists, 40 from each arm. The two control teams that did not attend are coded as raising zero. Funding is the total amount received from a team's jurors, in rupees.

TABLE A.6—PROBLEM SOLVING: ROBUSTNESS TO THE DIFFICULTY CHOICE

Outcome coding	Treatment effect	SE	p	N
Difficulty-weighted score (main measure)	0.142**	(0.064)	0.029	4,388
Score residualized on the difficulty choice	0.140**	(0.063)	0.029	4,388
Unweighted score (first principal component)	0.136**	(0.064)	0.038	4,388

Notes: The table reports the treatment effect on problem solving under three codings of the outcome. In each of five Sudoku rounds, the student chose an easy or a hard puzzle; in round 1, the hard puzzle was assigned to a randomly selected half of students. The first row is the main measure, the average score across the five rounds with hard puzzles weighted 1.5 times. The second row removes the influence of the difficulty choice by residualizing the weighted score on the number of hard puzzles the student faced. The third row applies no difficulty weighting and uses the first principal component of the five round scores. All outcomes are standardized to the control group. Every regression controls for randomization strata, school infrastructure, individual covariates (age, household SES index, and computer literacy), and baseline fluid intelligence with an indicator for missing values. Standard errors, in parentheses, are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

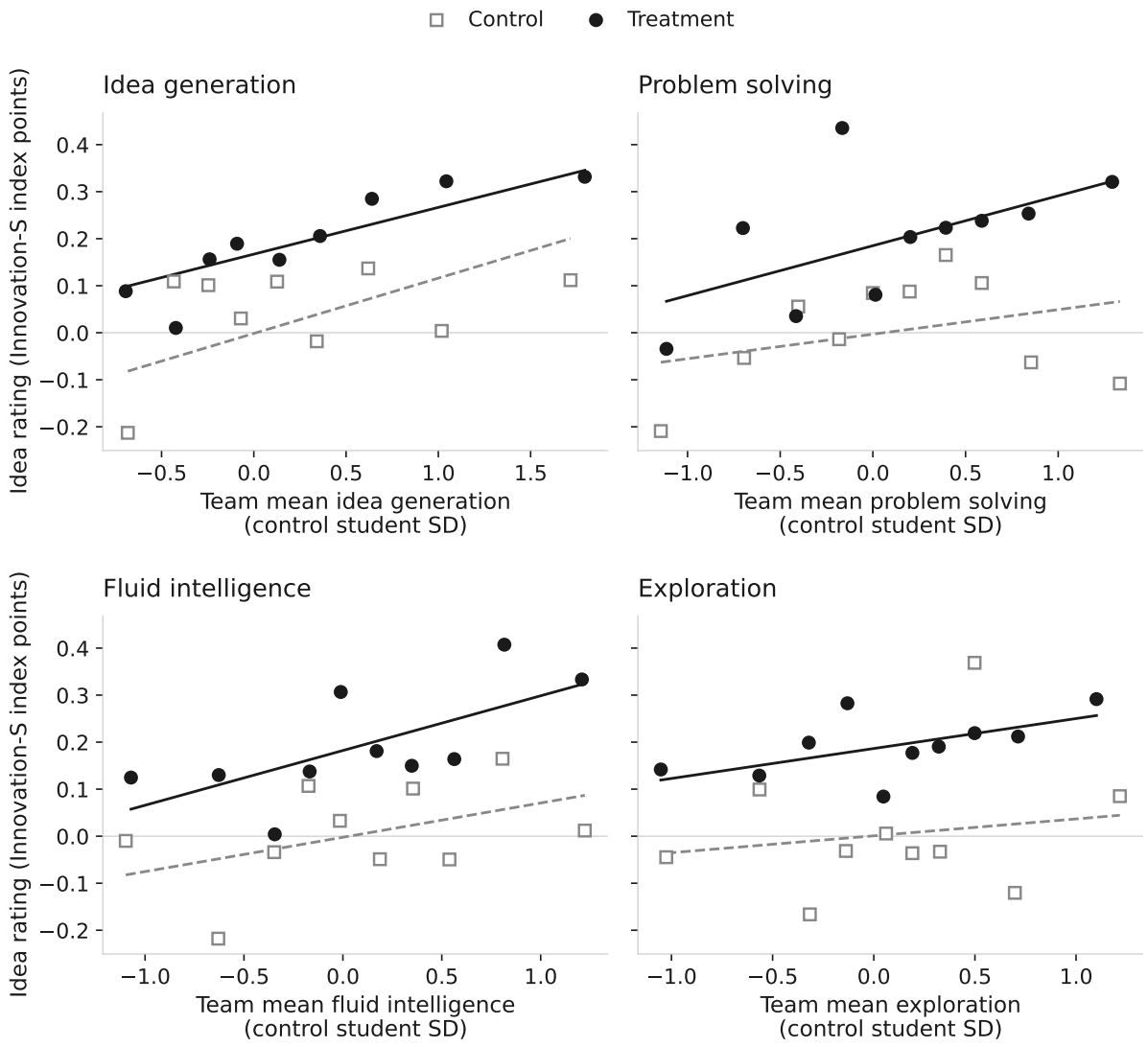


FIGURE A.3. IDEA QUALITY AND TEAM-AVERAGE INNOVATION-RELEVANT SKILLS

Notes: The sample comprises 568 control and 648 treatment teams with scored submissions and observed team-average skill measures. Idea quality is the Innovation-S index combining control-standardized novelty and value-add ratings. Team skill scores average the individually assessed scores of members with observed data; individual scores are standardized using control students. Points show mean idea quality against mean skill within pooled skill-score quantile bins, separately by treatment status. Bin boundaries are common across arms; there are ten bins per skill, except idea generation, where tied scores yield nine. Lines are fitted by ordinary least squares to individual teams, separately by treatment status. These associations are descriptive and conditional on submission.

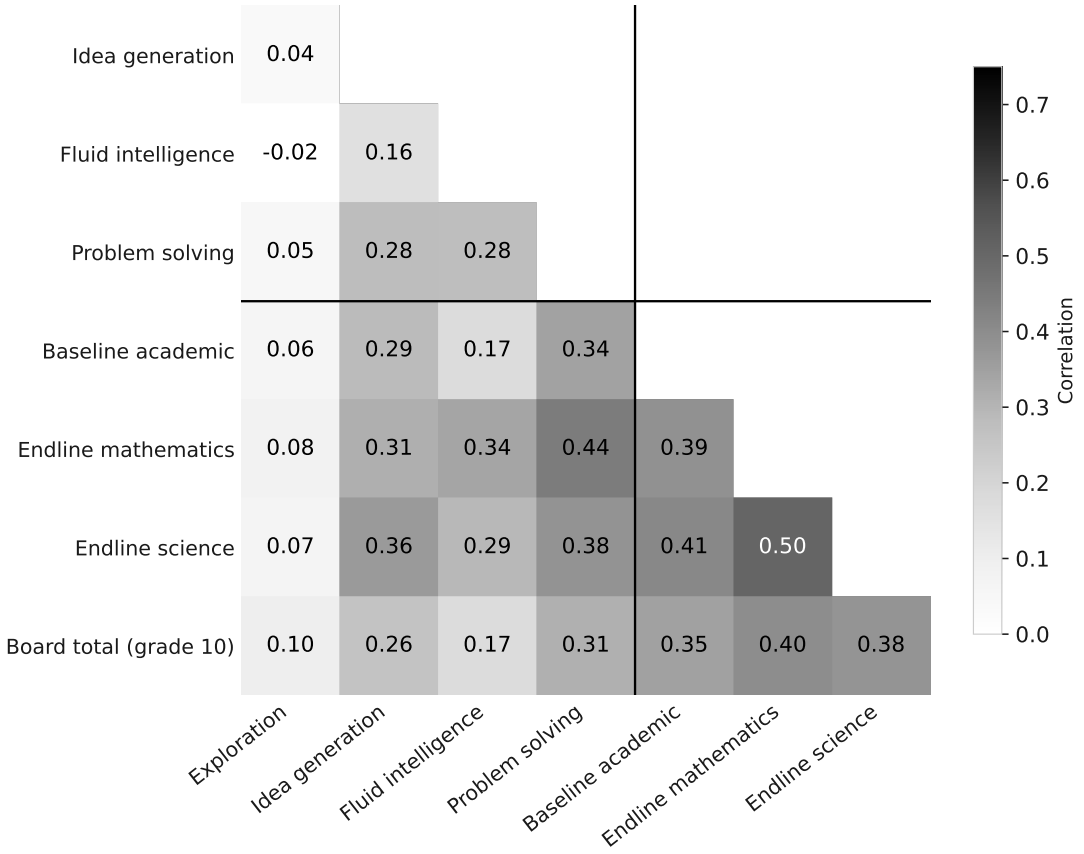


FIGURE A.4. CORRELATIONS AMONG SKILL AND ACADEMIC MEASURES

Notes: Pairwise correlations among the four endline skill measures and four academic measures, computed over the 1,600 control students with a non-imputed baseline academic score and all eight measures observed. Each correlation appears once; darker cells indicate stronger correlations, and the solid lines separate the skill block from the academic block. Appendix Table D.2 reports Cronbach's α for the skill measures.

TABLE A.7—HETEROGENEOUS TREATMENT EFFECTS

	By school gender composition			By baseline academic score			
	Boys' schools	Girls' schools	Difference p	Below median	Above median	Difference p	N
Panel A. Innovation Challenge production (Table 4 outcomes)							
Participation (pp)	12.9***	3.7	0.031	10.2***	5.5**	0.107	4,388
	(3.1)	(2.8)		(2.5)	(2.6)		
Novelty (SD)	0.241***	0.140**	0.364	0.205***	0.168**	0.716	1,493
	(0.090)	(0.064)		(0.070)	(0.079)		
Value-add (SD)	0.222**	0.079	0.188	0.174**	0.116	0.563	1,493
	(0.086)	(0.063)		(0.068)	(0.077)		
Feasibility (SD)	0.180**	0.047	0.157	0.160**	0.054	0.303	1,493
	(0.076)	(0.056)		(0.069)	(0.069)		
Willingness to fund (SD)	0.236**	0.138*	0.440	0.204***	0.163*	0.719	1,493
	(0.097)	(0.080)		(0.076)	(0.091)		
Panel B. Innovation-relevant skills (Table 5 outcomes)							
Exploration index (points)	0.052	0.130***	0.130	0.091***	0.097***	0.890	4,388
	(0.035)	(0.036)		(0.031)	(0.035)		
Idea generation (SD)	0.096	0.503***	0.011	0.200***	0.456***	0.014	4,388
	(0.087)	(0.123)		(0.072)	(0.102)		
Idea evaluation (SD)	-0.003	0.230**	0.082	0.058	0.203**	0.089	4,388
	(0.082)	(0.100)		(0.063)	(0.083)		
Written expression (SD)	-0.007	0.081	0.504	0.003	0.090	0.264	4,388
	(0.094)	(0.090)		(0.071)	(0.069)		
Fluid intelligence (SD)	0.062	0.153**	0.339	0.020	0.222***	0.014	4,388
	(0.069)	(0.069)		(0.063)	(0.063)		
Problem solving (SD)	0.250**	0.046	0.164	0.158**	0.126	0.697	4,388
	(0.115)	(0.076)		(0.070)	(0.083)		
Emotional intelligence (SD)	0.021	0.039	0.848	-0.021	0.094	0.136	4,388
	(0.066)	(0.069)		(0.059)	(0.062)		
Challenge-seeking (SD)	0.054	-0.011	0.686	0.019	0.019	0.998	4,388
	(0.096)	(0.123)		(0.083)	(0.104)		
Panel C. Academic outcomes (Table 6 outcomes)							
Endline mathematics (SD)	-0.044	0.025	0.615	-0.112	0.118	0.009	4,388
	(0.097)	(0.094)		(0.069)	(0.087)		
Endline science (SD)	-0.034	0.079	0.358	-0.011	0.075	0.334	4,388
	(0.085)	(0.086)		(0.059)	(0.088)		
Grade 10 total score (SD)	0.014	0.242*	0.326	0.158	0.108	0.681	5,633
	(0.187)	(0.138)		(0.141)	(0.116)		

Notes: Participation and skill outcomes use 4,388 endline-assessed students; team ratings use 1,493 team units; the grade 10 row uses 5,633 students. Outcomes and units follow Tables 4–6. The first two columns of each block report subgroup effects. Stars test each effect against zero; the "Difference" columns report p-values for equality of the two effects.

Student and team-rating rows use interacted versions of equations (2) and (3), respectively, adding the group indicator and its interaction with treatment. The girls-only indicator is removed from the strata controls when it defines the groups. Individual academic groups use the control-group median of the combined 27-item baseline academic score among students with observed scores; missing scores are imputed as in Section V. Team-rating groups use the median of team-average baseline academic scores among control teams, recomputed within each grouping. The remaining controls and baseline measures follow the corresponding results-table row. Standard errors, in parentheses, are clustered by school and use the full variance-covariance matrix for each subgroup effect; for team ratings, variances also incorporate between-grouping variation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A.8—HETEROGENEOUS TREATMENT EFFECTS: MANAGEMENT AND LOCATION

	By management			By location			
	Tribal (ST)	Social (SC)	p	Urban	Rural	p	N
Panel A. Innovation Challenge production (Table 4 outcomes)							
Participation (pp)	11.9***	4.6	0.109	3.6	10.5***	0.161	4,388
	(3.3)	(2.8)		(4.1)	(2.4)		
Novelty (SD)	0.274***	0.109	0.143	0.038	0.273***	0.050	1,493
	(0.076)	(0.078)		(0.093)	(0.068)		
Value-add (SD)	0.228***	0.072	0.152	-0.022	0.242***	0.021	1,493
	(0.077)	(0.073)		(0.087)	(0.067)		
Feasibility (SD)	0.185***	0.041	0.138	-0.034	0.191***	0.026	1,493
	(0.068)	(0.065)		(0.073)	(0.062)		
Willingness to fund (SD)	0.287***	0.091	0.105	0.076	0.246***	0.219	1,493
	(0.091)	(0.080)		(0.102)	(0.083)		
Panel B. Innovation-relevant skills (Table 5 outcomes)							
Exploration index (points)	0.092**	0.095***	0.961	0.091**	0.095***	0.942	4,388
	(0.040)	(0.030)		(0.037)	(0.033)		
Idea generation (SD)	0.149	0.460***	0.057	0.349**	0.294***	0.766	4,388
	(0.103)	(0.118)		(0.159)	(0.086)		
Idea evaluation (SD)	-0.009	0.236**	0.076	0.099	0.133*	0.819	4,388
	(0.096)	(0.092)		(0.124)	(0.078)		
Written expression (SD)	-0.115	0.178*	0.032	0.032	0.045	0.934	4,388
	(0.096)	(0.089)		(0.125)	(0.082)		
Fluid intelligence (SD)	-0.004	0.213***	0.050	0.104	0.115*	0.934	4,388
	(0.073)	(0.074)		(0.093)	(0.067)		
Problem solving (SD)	0.088	0.189**	0.469	0.017	0.210***	0.185	4,388
	(0.095)	(0.094)		(0.120)	(0.076)		
Emotional intelligence (SD)	-0.051	0.103	0.108	0.070	0.009	0.604	4,388
	(0.066)	(0.068)		(0.094)	(0.061)		
Challenge-seeking (SD)	-0.015	0.049	0.685	0.041	0.007	0.831	4,388
	(0.111)	(0.110)		(0.131)	(0.094)		
Panel C. Academic outcomes (Table 6 outcomes)							
Endline mathematics (SD)	-0.175**	0.142	0.019	0.142	-0.089	0.184	4,388
	(0.087)	(0.097)		(0.131)	(0.088)		
Endline science (SD)	-0.045	0.090	0.261	0.037	0.021	0.910	4,388
	(0.075)	(0.093)		(0.107)	(0.080)		
Grade 10 total score (SD)	-0.011	0.266	0.239	0.101	0.156	0.827	5,633
	(0.157)	(0.168)		(0.200)	(0.142)		

Notes: Specification, samples, and inference follow Appendix Table A.7 exactly; the moderating characteristic is management (reference: tribal-welfare) in the first block and location (reference: urban) in the second, each removed from the strata controls and entered directly with its interaction. Standard errors, in parentheses, are clustered at the school level and, for idea ratings, incorporate between-grouping variation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A.9—GENDER AND BASELINE-SCORE INTERACTIONS: JOINT AND WITHIN-SUBSAMPLE ESTIMATES

Interaction	Coefficient	SE	p	N
Idea generation: $T\times$ Girls' school	0.337**	(0.142)	0.020	4,388
Idea generation: $T\times$ Above-median score	0.198**	(0.099)	0.049	4,388
Fluid intelligence: $T\times$ Girls' school	0.048	(0.092)	0.605	4,388
Fluid intelligence: $T\times$ Above-median score	0.194**	(0.081)	0.018	4,388
Idea generation: $T\times$ Girls' school, SC only	0.619***	(0.201)	0.004	2,298
Idea generation: $T\times$ Girls' school, ST only	0.050	(0.139)	0.722	2,090

Notes: The first four rows jointly estimate the two interactions in one regression per outcome, $Y=\beta_0+\beta_1 T+\beta_2 G+\beta_3(T\times G)+\beta_4 A+\beta_5(T\times A)+\mathbf{X}\gamma+\varepsilon$, where G indicates a girls-only school and A indicates an above-median baseline academic score. The last two rows estimate the Table A.7 gender interaction separately within the 40 social-welfare (SC) and the 40 tribal-welfare (ST) schools; gender composition and management are fully crossed in the design, 20 schools per cell. All specifications otherwise follow Table A.7, including each outcome's baseline control; standard errors are clustered at the school level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

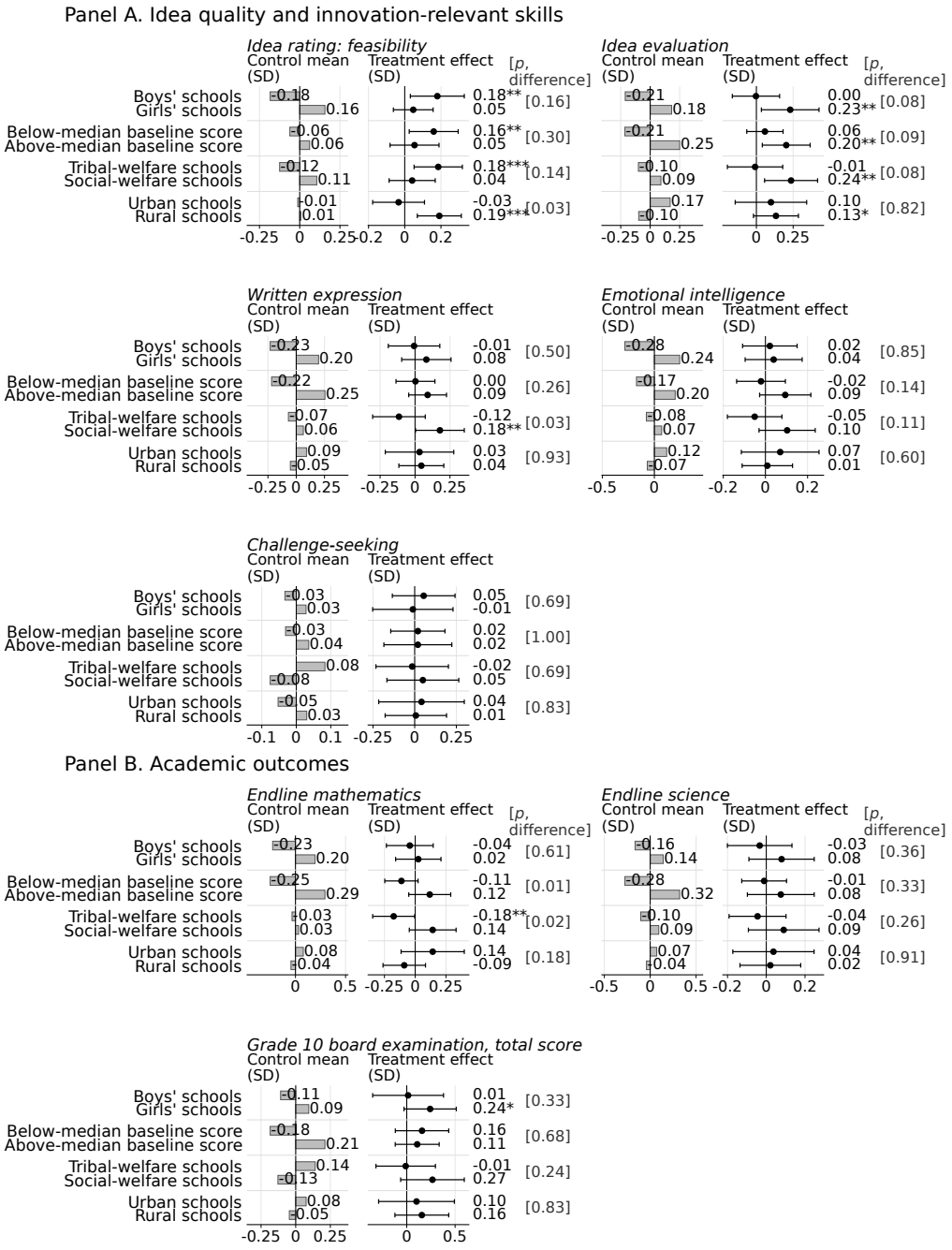


FIGURE A.5. TREATMENT EFFECTS BY SUBGROUP: ADDITIONAL OUTCOMES

Notes: Bars show control-group means; points show treatment effects and 95 percent confidence intervals for eight subgroups defined by school gender composition, baseline academic score, management, and location. The estimate is printed to the right of the interval, and the value in brackets is the p-value for equality of the two effects in the pair. The feasibility rating uses 1,493 teams, with teams that did not submit an idea assigned the value of no idea. The grade 10 total score uses 5,633 students with a board examination result. All other outcomes use the 4,388 students assessed at endline. Outcomes are in standard deviations of the control group. Baseline-score subgroups are defined by the control-group median of the baseline academic score for students and, for teams, of the team-average score among control teams, recomputed within each grouping of non-submitters. Effects are estimated from regressions that interact treatment with the subgroup indicator and include randomization strata, school infrastructure, individual or team controls, and the baseline measure of the outcome. Standard errors are clustered at the school level and, for feasibility, incorporate between-grouping variation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A.10—ALTERNATIVE SPECIFICATIONS: INNOVATION CHALLENGE PRODUCTION

Outcome	(1) No controls	(2) Strata indicators	(3) Strata + baseline	(4) Main specification	(5) Stratum-cell FE	N
Panel A. All teams						
Participation (pp)	7.5**	7.7***	7.8***	8.0***	8.0***	4,388
	(2.9)	(2.4)	(2.3)	(2.1)	(2.1)	
Novelty (SD)	0.229***	0.238***	0.242***	0.187***	0.187***	1,493
	(0.069)	(0.058)	(0.057)	(0.054)	(0.054)	
Value-add (SD)	0.190***	0.199***	0.200***	0.146***	0.146***	1,493
	(0.066)	(0.055)	(0.055)	(0.052)	(0.052)	
Feasibility (SD)	0.152**	0.161***	0.157***	0.109**	0.109**	1,493
	(0.062)	(0.049)	(0.050)	(0.047)	(0.046)	
Willingness to fund (SD)	0.205***	0.214***	0.219***	0.184***	0.183***	1,493
	(0.077)	(0.069)	(0.067)	(0.062)	(0.061)	
Panel B. Submitted teams only						
Novelty (SD)	0.226**	0.238***	0.255***	0.265***	0.264***	1,225
	(0.089)	(0.086)	(0.084)	(0.070)	(0.071)	
Value-add (SD)	0.162*	0.172**	0.185**	0.202***	0.200***	1,225
	(0.086)	(0.084)	(0.082)	(0.069)	(0.069)	
Feasibility (SD)	0.080	0.090	0.085	0.100	0.100	1,225
	(0.074)	(0.069)	(0.069)	(0.062)	(0.061)	
Willingness to fund (SD)	0.158*	0.168*	0.184**	0.206***	0.204***	1,225
	(0.092)	(0.089)	(0.086)	(0.073)	(0.072)	
Strata indicators	No	Yes	Yes	Yes	No	
Saturated stratum-cell fixed effects	No	No	No	No	Yes	
Baseline controls + missing flags	No	No	Yes	Yes	Yes	
School infrastructure controls	No	No	No	Yes	Yes	
Individual or team composition controls	No	No	No	Yes	Yes	

Notes: The table reports treatment effects on Innovation Challenge outcomes under five specifications. Panel A uses the full sample: participation is measured for the 4,388 students assessed at endline, in percentage points, and the four evaluator ratings are measured for 1,493 teams, comprising submitted teams and constructed teams of non-submitters. Panel B restricts the sample to the 1,225 teams that submitted a scored idea. Ratings are in standard deviations of the control group of the panel's sample. Column (1) regresses the outcome on treatment alone. Column (2) adds the three randomization-strata indicators (tribal-welfare management, girls-only composition, and rural location). Column (3) adds baseline controls with indicators for missing values: baseline mathematics and science scores for participation, and seven team-average baseline characteristics for the ratings. Column (4) is the main specification, which further adds school infrastructure controls and individual or team composition controls. Column (5) replaces the three strata indicators of column (4) with fixed effects for the eight strata cells. Idea-rating estimation and inference follow Section V. Because submission is itself affected by treatment, the estimates in Panel B are conditional on submission. Standard errors, in parentheses, are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A.11—ALTERNATIVE SPECIFICATIONS: INNOVATION-RELEVANT SKILLS

Outcome	(1) No controls	(2) Strata indicators	(3) Strata + baseline	(4) Main specification	(5) Stratum-cell FE	N
Panel A. Search behavior in the exploration task						
Exploration index (points)	0.088***	0.089***	0.092***	0.094***	0.096***	4,388
	(0.030)	(0.026)	(0.026)	(0.025)	(0.024)	
Initiation (pp)	5.6***	5.7***	5.7***	5.6***	5.7***	4,388
	(2.0)	(1.9)	(1.9)	(1.8)	(1.8)	
Breadth	0.138	0.144	0.147	0.161*	0.163*	4,388
	(0.105)	(0.094)	(0.094)	(0.092)	(0.088)	
Discovery	0.090**	0.091**	0.095**	0.101**	0.104***	4,388
	(0.044)	(0.042)	(0.041)	(0.039)	(0.038)	
Convergence	-0.252*	-0.255**	-0.265**	-0.256**	-0.262**	4,388
	(0.132)	(0.128)	(0.123)	(0.115)	(0.112)	
Panel B. Creativity						
Idea generation (SD)	0.310***	0.313***	0.316***	0.314***	0.314***	4,388
	(0.096)	(0.084)	(0.083)	(0.078)	(0.076)	
Idea evaluation (SD)	0.119	0.124*	0.126*	0.121*	0.120*	4,388
	(0.096)	(0.071)	(0.070)	(0.066)	(0.064)	
Written expression (SD)	0.031	0.036	0.039	0.040	0.041	4,388
	(0.095)	(0.075)	(0.074)	(0.065)	(0.063)	
Panel C. Analytical reasoning						
Fluid intelligence (SD)	0.108	0.110	0.113*	0.111**	0.109**	4,388
	(0.074)	(0.069)	(0.057)	(0.050)	(0.049)	
Problem solving (SD)	0.102	0.107	0.109	0.142**	0.141**	4,388
	(0.096)	(0.082)	(0.074)	(0.064)	(0.063)	
Panel D. Emotional intelligence and challenge-seeking						
Emotional intelligence (SD)	0.006	0.012	0.021	0.031	0.030	4,388
	(0.091)	(0.062)	(0.051)	(0.048)	(0.046)	
Challenge-seeking (SD)	0.009	0.009	0.015	0.019	0.022	4,388
	(0.083)	(0.081)	(0.081)	(0.078)	(0.072)	
Strata indicators	No	Yes	Yes	Yes	No	
Saturated stratum-cell fixed effects	No	No	No	No	Yes	
Baseline control + missing flag	No	No	Yes	Yes	Yes	
School infrastructure controls	No	No	No	Yes	Yes	
Individual controls	No	No	No	Yes	Yes	

Notes: The table re-estimates every Table 5 row under five specifications; outcome definitions, samples, and panel baselines are as there. Column (1) regresses the outcome on treatment alone; column (2) adds the three randomization-strata indicators (tribal-welfare management, girls-only composition, rural location); column (3) adds the panel's baseline control with its missing-value flag (the baseline exploration index for Panels A and B, baseline fluid intelligence for Panel C, and, in Panel D, the baseline eyes test for emotional intelligence and baseline conscientiousness for challenge-seeking); column (4) repeats the main specification, adding school infrastructure and individual controls (age, household assets, computer skills), and reproduces it exactly; column (5) keeps the main specification but replaces the three strata indicators with the eight saturated stratum-cell fixed effects formed by their interaction. Initiation is reported in percentage points; Breadth, Discovery, and Convergence are in natural units. The exploration index is the equal-weighted mean of the four control-standardized components, with Convergence reverse-signed. Outcomes in Panels B–D are standardized to have control-group mean zero and standard deviation one. Sample sizes are identical across the five columns within each row. Standard errors, in parentheses, are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

TABLE A.12—ALTERNATIVE SPECIFICATIONS: ACADEMIC OUTCOMES

Outcome	(1) No controls	(2) Strata indicators	(3) Strata + baseline	(4) Main specification	(5) Stratum-cell FE	N
Panel A. Endline assessments						
Mathematics	-0.044	-0.039	-0.022	-0.007	-0.010	4,388
	(0.110)	(0.095)	(0.078)	(0.067)	(0.063)	
Science	0.008	0.012	0.018	0.027	0.026	4,388
	(0.095)	(0.082)	(0.066)	(0.061)	(0.058)	
Panel B. Grade 10 board examinations						
Mathematics	0.104	0.106	0.112	0.112	0.107	5,632
	(0.157)	(0.155)	(0.156)	(0.147)	(0.143)	
Science	-0.013	-0.010	-0.008	-0.006	-0.012	5,632
	(0.133)	(0.130)	(0.131)	(0.119)	(0.113)	
Social studies	0.001	0.004	0.010	0.023	0.015	5,632
	(0.109)	(0.102)	(0.104)	(0.093)	(0.089)	
First language	0.115	0.119	0.130	0.149*	0.145*	5,632
	(0.106)	(0.094)	(0.094)	(0.088)	(0.086)	
Second language	0.172	0.175	0.182	0.199*	0.196*	5,631
	(0.111)	(0.108)	(0.110)	(0.106)	(0.104)	
Third language	0.126	0.126	0.135	0.153*	0.153*	5,631
	(0.098)	(0.089)	(0.092)	(0.086)	(0.084)	
Total score	0.112	0.116	0.126	0.135	0.130	5,633
	(0.125)	(0.120)	(0.123)	(0.115)	(0.111)	
Passed (pp)	0.53	0.54	0.59	0.64	0.61	5,642
	(0.91)	(0.90)	(0.89)	(0.83)	(0.77)	
Strata indicators	No	Yes	Yes	Yes	No	
Saturated stratum-cell fixed effects	No	No	No	No	Yes	
Baseline controls + missing flags	No	No	Yes	Yes	Yes	
School infrastructure controls	No	No	No	Yes	Yes	
Individual controls	No	No	No	Yes	Yes	

Notes: The table re-estimates the Table 6 academic rows under five specifications; outcome definitions and samples are as there, and the behavioral rows (homework share, attendance) are not part of the specification ladder. Column (1) regresses the outcome on treatment alone; column (2) adds the three randomization-strata indicators (tribal-welfare management, girls-only composition, rural location); column (3) adds the row's baseline control(s) with missing-value flags (baseline mathematics for mathematics, baseline science for science, and both for the remaining grade 10 rows); column (4) repeats the main specification, adding school infrastructure and individual controls (age, household assets, computer skills), and reproduces it exactly; column (5) keeps the main specification but replaces the three strata indicators with the eight saturated stratum-cell fixed effects formed by their interaction. Test-score outcomes are standardized to the control group; Passed is in percentage points with absentee and withheld results coded as not passing. Sample sizes are identical across the five columns within each row. Standard errors, in parentheses, are clustered at the school level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

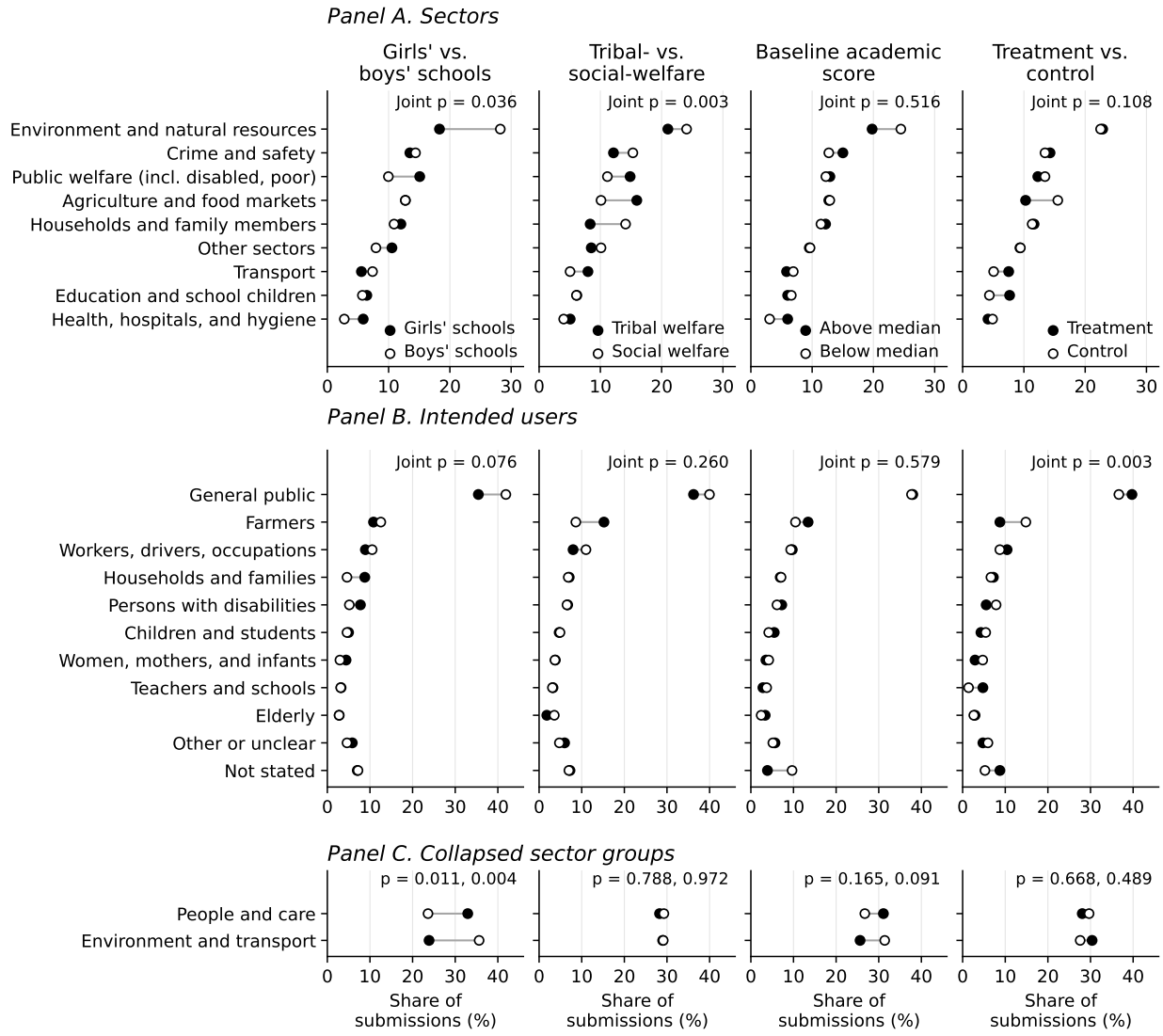


FIGURE A.6. DIRECTION OF IDEAS: SECTORS AND INTENDED USERS BY TEAM ATTRIBUTES

Notes: The figure reports what the 1,225 scored Innovation Challenge submissions set out to solve (Panel A, the sector the team selected on the submission form) and for whom (Panel B, intended users classified from the submission's free-text description by the keyword rules in the replication code; 5.3 percent are unmatched and appear as Other or unclear), by four team attributes. Filled and open circles are the two groups' shares of submissions, the gray segment is their gap, and categories are ordered by the pooled share. The columns compare girls' and boys' schools (683 and 542 submissions), tribal-welfare and social-welfare schools (552 and 673), teams above and below the median baseline academic score of control submitted teams (566 and 621; 38 teams with no matched baseline member are excluded), and treatment and control schools (652 and 573). Panel C collapses the sectors into two groups, people and care (public welfare; health, hospitals, and hygiene; households and family members) and environment and transport, defined after inspecting Panel A. Each panel reports the *p*-value of a joint test that the category shares are equal across the two groups, with standard errors clustered at the school level.

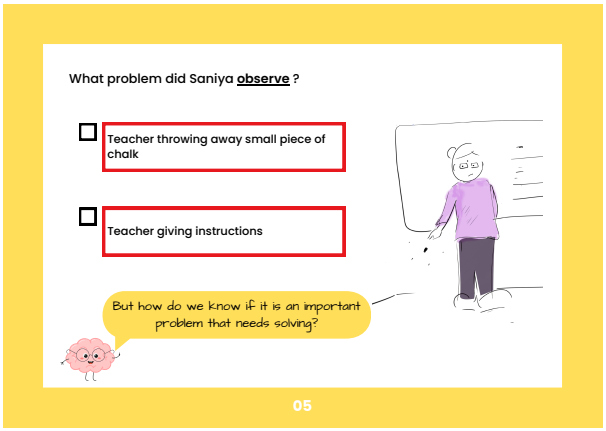
B. The Curriculum: Content and Delivery

TABLE B.1—THE THINK & MAKE CURRICULUM AND DELIVERY MODEL

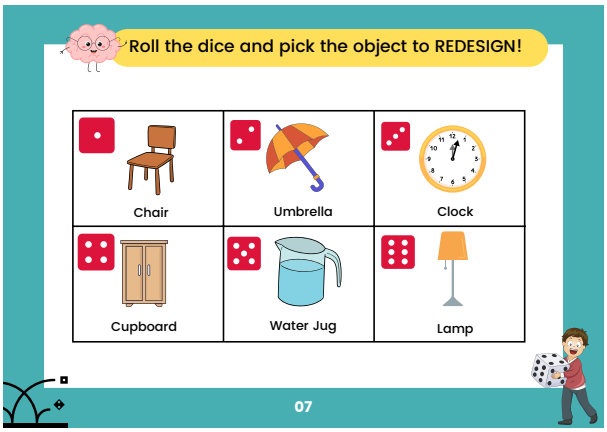
Study year and block	Instructional progression	What students do	Required work product
Panel A: Curriculum Sequence and Student Work			
Year 1, Block 1: Planet	Guided application of the problem-solving cycle to environmental problems	Observe resource use; identify and select a problem; interview affected people; investigate causes and effects; generate and evaluate solutions	Problem inventory; interview notes; cause-and-effect analysis; idea-selection rubric; prototype; presentation
Year 1, Block 2: People	Guided application of the same cycle to problems affecting people's quality of life	Identify problems through observation and interviews; investigate a selected problem; generate alternatives; choose a solution	Problem statement; interview evidence; written problem story; idea-selection rubric; prototype; presentation
Year 2, Block 1: Design Evolutions	Worked examples, followed by constrained redesign and independent problem solving	Study how everyday objects evolved; solve a guided story-card problem; redesign an assigned object for a selected user; identify and score problems and ideas from students' own surroundings	Story-card solution; mock-up and sketch prototypes; problem- and idea-ranking sheets; resource plan; final prototype and presentation
Year 2, Block 2: Smarter Everyday	Technical experimentation, followed by guided application and independent technology integration	Build input–sensor–output circuits; experiment with sensors and outputs; assess when sensors are useful; solve guided sensor problems; develop and evaluate their own applications	Circuit exercises; sensor-use critiques; sensor-application ideas; storyboard, sketch, or working mock-up; resource plan; final presentation
Panel B: Delivery Model			
Student organization	Students worked primarily in fixed teams of four. Each team was paired with a co-team for structured feedback.		
Student leaders	Two trained student leaders facilitated each classroom, divided facilitation and observation responsibilities, managed time and materials, and guided teams through the workbook instructions.		
Teacher role	A teacher was expected to be present to support classroom management, mediate problems, advise teams when needed, and help organize showcases, but did not deliver the curriculum.		
Curriculum materials	Step-by-step student workbooks, story and activity cards, presentation cards, problem- and idea-selection rubrics, and team portfolios.		
Prototyping materials	A classroom Curiosity Box containing low-cost construction materials; Year 2 added batteries, circuit boards, sensors, LEDs, motors, propellers, and buzzers.		
Feedback and iteration	Co-team feedback, whole-class feedback, user testing, individual reflection, team reflection, and cross-session improvement prompts.		
Scheduling and culmination	Seven to nine weekly 90-minute sessions per block outside regular academic instruction, ending with team presentations of the problem, evidence, selected solution, prototype, and feedback.		

Notes: Panel A summarizes the four curriculum blocks delivered during the two study years. Each block required teams to move from problem identification to investigation, idea generation and selection, prototyping, testing, feedback, and presentation, with the degree of scaffolding declining over time. Panel B summarizes the classroom delivery protocol. Sources: Inqui-Lab student workbooks, student-leader handbook, team portfolios, session-detailing document, and Product Note.

Panel A: Guided



Panel B: Semi-open



Panel C: Independent

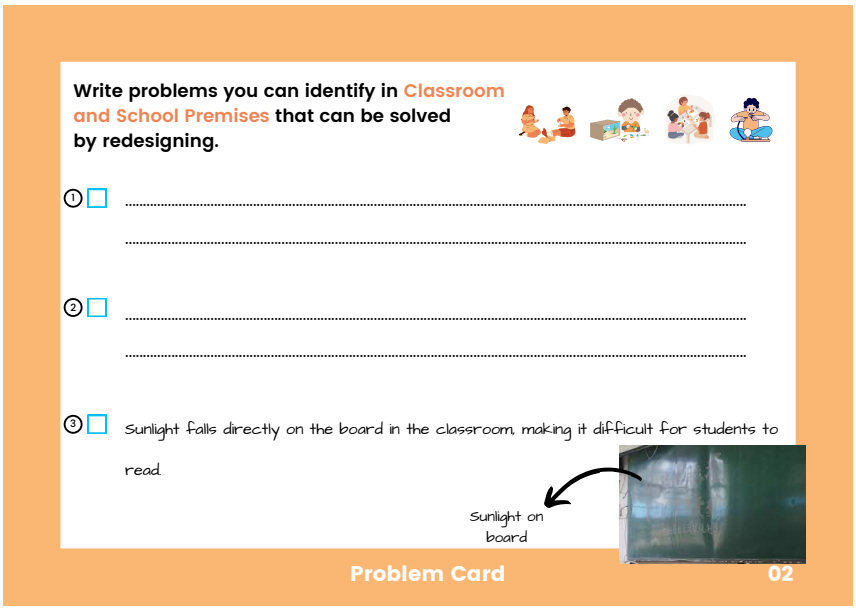


FIGURE B.1. PROGRESSIVE RELEASE OF AUTONOMY IN THINK & MAKE

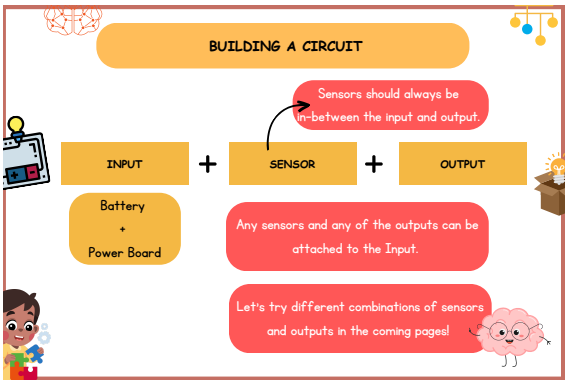
Notes: Panel A shows a guided story-card comprehension task: the problem and possible interpretations are supplied. Panel B shows a semi-open design challenge: the object is assigned by a die roll and the user is specified, while the redesign is generated by the team. Panel C shows independent problem discovery: teams identify problems in their own school environment before ranking problems, generating and rating ideas, and planning a prototype. Across the three stages, students gain greater responsibility for defining the problem and solution; the progression involves more than a reduction in the amount of written guidance. Source: Inqui-Lab Foundation student workbooks.



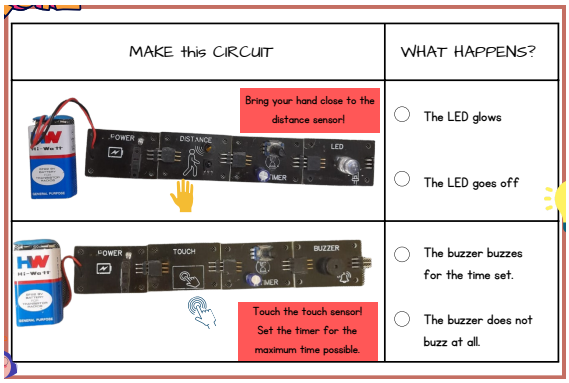
FIGURE B.2. INDEPENDENT PROBLEM DISCOVERY: FROM TWELVE PROBLEMS TO ONE PROTOTYPE

Notes: The panels reproduce pages from the Year 2, Block 1 (Design Evolutions) student workbook for the independent stage, in the order teams are asked to complete them. Panel A is one of four problem cards asking teams to list problems they could solve by redesign, three per setting (classroom and school, home and everyday objects, public places, and other), twelve in all. Panel B asks teams to shortlist five problems and score each on the number of people affected, how interesting it is, and their confidence in solving it, circling the highest total. Panel C shows two of the four spaces in which teams are asked to write and draw ideas for the selected problem. Panel D asks teams to score each idea on whether it is new, solves the problem fully, and is easy to prototype, circling the highest total. Panel E asks teams to choose between this idea and an idea from an earlier session. Panel F asks teams to list the materials for the prototype from the classroom kit and elsewhere. Source: Inqui-Lab Foundation student workbook, Block 1: Design Evolutions, PDF pages 150, 152, 153, 155, 156, and 158.

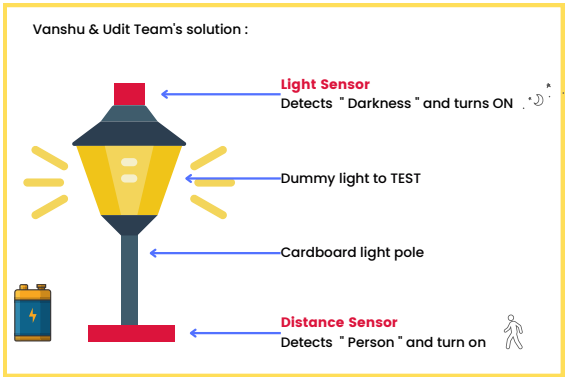
Panel A: Basic circuit components



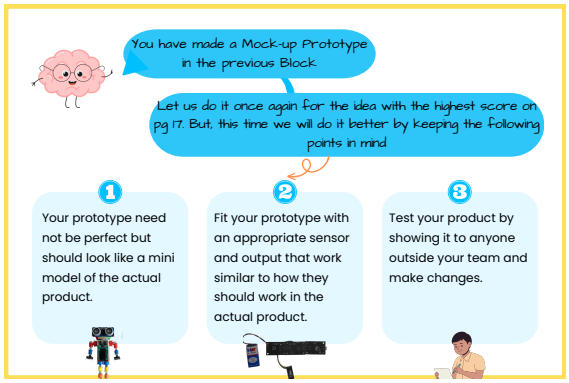
Panel B: Structured experimentation



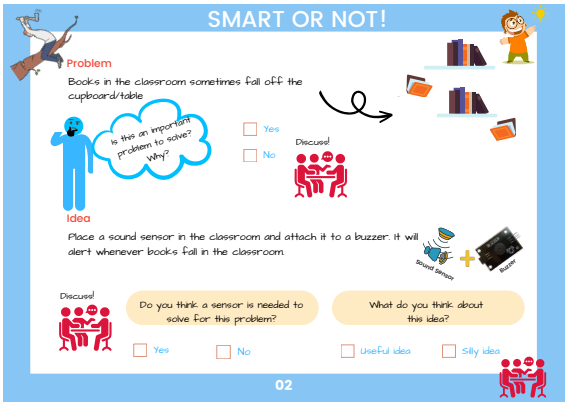
Panel C: Guided sensor solution



Panel D: Working-prototype requirement



Panel E: Judging when a sensor helps



Panel F: Independent sensor application

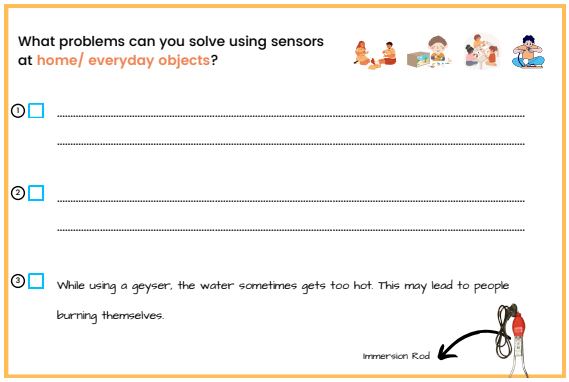


FIGURE B.3. LEARNING TO BUILD AND USE SENSORS IN SMARTER EVERYDAY

Notes: The panels reproduce selected pages from the Year 2, Block 2 (Smarter Everyday) student workbook in sequence. Panel A introduces the basic circuit components: power supply, sensor, and output. Panel B asks teams to assemble circuits and record what happens. Panel C presents a suggested street-light solution using light and distance sensors. Panel D asks teams to build a functioning prototype, test it with someone outside the team, and revise it. Panel E asks teams to assess whether a problem is important and whether it requires a sensor. Panel F asks teams to identify problems at home that sensors could help solve. Source: Inqui-Lab Foundation, Block 2: Smarter Everyday, PDF pages 11, 17, 46, 52, 146, and 164.

C. The Innovation Challenge: Measurement and Materials

TABLE C.1—INNOVATION-S: ITEMS, CONSTRUCTION, AND VALIDATION

Item	Dimension	Adult-survey mean	Student mean	Inter-rater r
Panel A: items				
Offers a new solution to a problem	Novelty	4.34	3.64	0.26
Improves on an existing solution	Novelty	3.93	3.92	0.26
Combines existing solutions	Novelty	4.07	3.74	0.27
Transplants a solution across contexts	Novelty	3.71	3.68	0.24
Solves a problem affecting many people	Value-add	4.48	3.92	0.23
Is unconventional compared with other solutions	Value-add	4.07	3.52	0.19
Will improve the lives of those who use it	Value-add	4.55	4.13	0.23
Panel B: scale statistics				
Adult-survey respondents	44 answered at least one item; 40 complete on all seven			
Adult-survey Cronbach's α (complete cases)	0.485			
Student-sample scored submissions (two evaluators)	1,189			
Student-sample Cronbach's α	0.950			
First principal component, share of item variance	0.775			
Inter-rater correlation of the seven-item index	0.300 ($N = 1,181$ submissions scored by both evaluators)			
Treatment effect on the index, evaluator 1's scores only (SD)	0.225 (SE 0.066, $p = 0.001$)			
Treatment effect on the index, evaluator 2's scores only (SD)	0.179 (SE 0.065, $p = 0.007$)			

Notes: The items were developed from interviews with 14 adult innovators and a survey of 44 adults involved in developing products. The scale contains four novelty and three value-add items, each rated 1–5. Adult-survey means use respondents answering each item. Student means use submissions with complete mean item scores; inter-rater correlations use submissions with complete scores from both evaluators. Cronbach's α measures internal consistency. For submissions with two ratings, the index averages the two evaluators' scores after standardizing within evaluator. The final two rows estimate treatment effects separately using each evaluator's scores among submitted teams, with the team-level specification in equation (3) and school-clustered standard errors. Although the two evaluators' scores are weakly correlated (0.30), the estimated effect is similar under either evaluator's scores (0.225 and 0.179 standard deviations). Thirty-six submissions were omitted from the set of submissions provided to the evaluators. Two large language models (ChatGPT, OpenAI; Claude, Anthropic) scored these submissions independently on the same items, using each team's drawing and a machine translation of its audio recording, without information on treatment status. Their scores are averaged in the same way as the two evaluators' scores. Excluding these submissions leaves the estimates in Table 4 essentially unchanged (0.186, 0.141, 0.111, and 0.174 standard deviations).

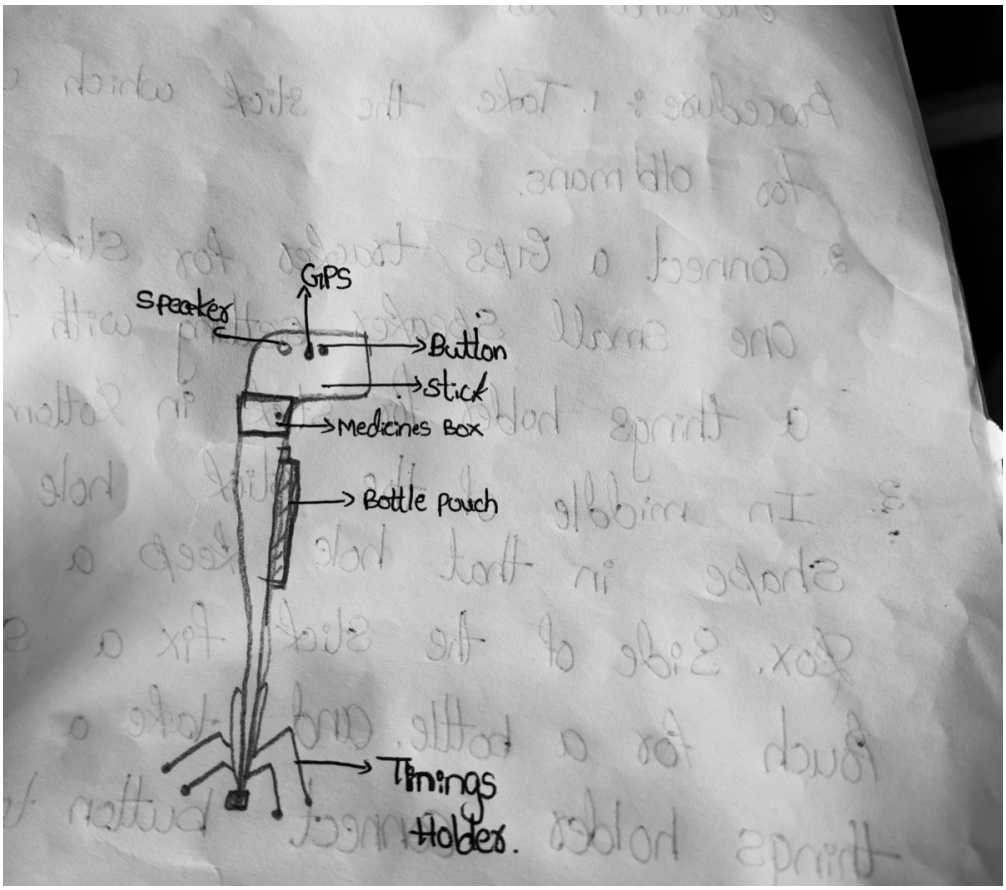
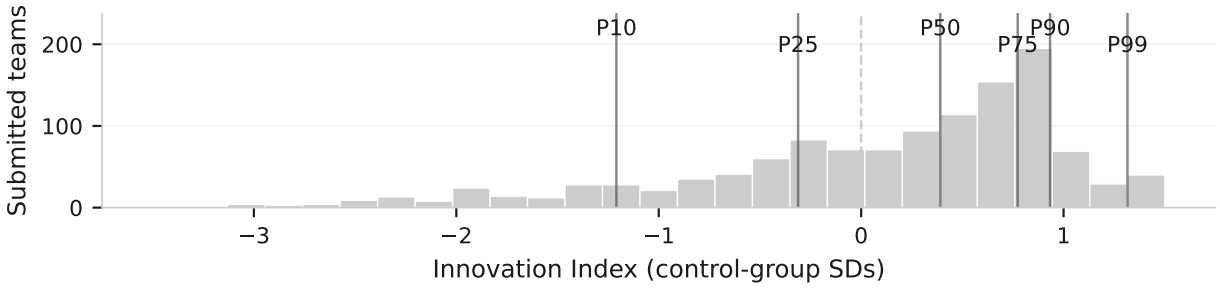


FIGURE C.1. A SUBMITTED IDEA DRAWING

Notes: Drawing submitted by a control-school team in the Year 1 Innovation Challenge. The proposed walking stick is designed for elderly people living alone and includes a GPS tracker, a speaker for medication reminders, a medicines box, and a bottle pouch. The paper analyzes Year 2 submissions, which used the same drawing format. The faint mirrored text is the team's written procedure on the reverse side of the page.



P10: index -1.21 SD

Problem: It is difficult for some companies to publicizr their brand successfully. They waste so much money on this.

Solution: To successfully publicize their brand, a helicopter is used. The brand is promoted via a helicopter with a speaker. As it goes around the town, a poster is also suspended so that people can look at it and...

P25: index -0.31 SD

Problem: how to be saved from monkeys, when we are classes, dorms, washrooms, we dont know if there is amonkey there or not, when we...

Solution: so we thought of a sensor which can be used anywhere for exampe dorm door or washroom door, if there is a monkey coming then the sensor buzzes and we will get intimitaded, because of the sound we will be...

P50: index 0.39 SD

Problem: Environment and natural resource: Water wastage due tonegligence towards motors. Often people forget to switch their water...

Solution: censor in water tanks and buzzers to indicate when the motor is full

P75: index 0.77 SD

Problem: Our problem is borewells. Many children are dying as they are falling into borewells while playing near them.

Solution: Our solution is to build it high and build steps in it along with placing sensors. The sensor will light up in blue and red colors in forests and villages respectively. This will alert the kids and blind...

P90: index 0.93 SD

Problem: when we use water, we release into drainage, but there are plastics, different types of toxics, other materials are also being...

Solution: so instead of directly releasing it , we can use this process where we recycle water and release them, we use 5 tanks, and there are different types of cleaning processes, in the first tank we use a process...

P99: index 1.32 SD

Problem: Blind, uneducated and old people find it difficult to access while travelling through a bus. They often get confused regarding...

Solution: When one goes to the bus station, the team proposes a sensor on each bus. This sensor gets activated once someone touches the stand on the bus and let people know where this it is headed. Additionally,...

FIGURE C.2. SUBMITTED IDEAS AT FIXED PERCENTILES OF RATED IDEA QUALITY

Notes: The histogram shows the distribution of evaluator-rated idea quality across submitted teams, measured by the combined novelty and value-add index in control-group standard deviations (Section IV.A). The six example submissions sit at the 10th, 25th, 50th, 75th, 90th, and 99th percentiles of that distribution; at each percentile the example is the submitted team with the most complete written record within a band around the percentile, selected without using treatment status, which is not shown. Each card reproduces, as submitted, the team's stated problem and proposed solution, together with the idea's index value; text beyond the card's space is truncated. The scale's construction and validation are reported in Appendix Table C.1.

Panel A. Tablet assessment: girls' school



Panel B. Tablet assessment: boys' school



Panel C. Geometry-box prize for the search task



Panel D. A finalist team presents its idea



Panel E. A team demonstrates its prototype



FIGURE C.3. DATA COLLECTION AND THE SHARK TANK FINALS

Notes: Panels A and B show individual tablet assessments in girls' and boys' schools. Panel C shows a student receiving a geometry box for the highest profit in the tablet search task. Panels D and E show finalist teams presenting ideas and prototypes to jurors at the Shark Tank finals in Hyderabad, Telangana. Faces are blurred to protect participants' privacy.

D. Skill Tasks: Design and Scoring

TABLE D.1—DESIGN AND PAYOFF STRUCTURE OF THE INCENTIVIZED SEARCH TASK

	Fresh juice	Pani puri	Ice cream
Panel A: stand parameters			
Maximum hourly profit (V_s, ₹)	120	180	250
Penalty per unit of distance (λ_s, ₹)	10	20	30
Optimal location	Market	School	Homes
Optimal working hours	13	12	14
Optimal menu items	6	8	7
Optimal selling price (₹)	15	13	14
Share of configurations with zero profit	0.257	0.540	0.633
Panel B: game constants			
Rounds	12		
Choice space	3 stands × 3 locations × 10 hours × 10 menu sizes × 10 prices = 9,000		
Feedback per round	own profit; choices and profits from completed rounds		
Worked example (round 1)	fresh juice at the market, 5 items, 9 hours, ₹19: ₹30 per hour		
Display from round 2	student's best configuration and profit so far		
Input selection	round 1: active selection; rounds 2–12: previous choices pre-filled		
Prize	a geometry box for the highest single-round hourly profit in the class		

Notes: Panel A reports each stand's payoff parameters in the units students faced. Profit in a round equals the stand's maximum hourly profit minus the penalty rate times the sum of absolute deviations between the chosen inputs and the optimal inputs across the four dimensions (location, working hours, menu items, selling price), floored at zero (equation 1). Working hours, menu items, and price each have a 10-point choice space, and location has a 3-point choice space; a one-step deviation costs the same on every dimension. The zero-profit share is the fraction of the stand's 3,000 configurations returning zero profit and is computed from the payoff function. None of these parameters was disclosed to students. Panel B reports the game constants. Payoffs are deterministic: in the data, repeated plays of the same configuration always return the same profit.

TABLE D.2—SKILL BATTERY: ADMINISTRATION AND PSYCHOMETRICS

	Eyes test (RMET)	Fluid intelligence (ICAR)	Creativity tasks (PISA)	Sudoku task
Measured outcome	Emotional intelligence	Fluid intelligence	Idea generation; idea evaluation; written expression	Problem solving; challenge-seeking
Source design	Reading the Mind in the Eyes Test (Baron-Cohen et al. 2001)	International Cognitive Ability Resource (Condon and Revelle 2014)	PISA creative-thinking tasks (OECD 2024)	Incentivized easy-versus-hard choice (Alan, Boneva, and Ertac 2019), Sudoku in place of number-matching
Items or rounds	11 items	12 items	3 written tasks	5 rounds, four-by-four puzzles
Response format	4 labeled options, English and Telugu	4 or 6 options, depending on the item	Open-ended text, English or Telugu	Easy-or-hard choice, then the puzzle; 2 minutes per round
Scoring	Two-parameter item response model	Two-parameter item response model	Task-specific rubrics applied by a large language model	Difficulty-weighted average score (hard $\times 1.5$); challenge-seeking = share of hard choices, rounds 2–5
Cronbach's α	0.46	0.79	0.71 (idea generation, three idea scores)	0.84
Baseline form	Longer form, shortened to the most discriminating items	10-item form, recalibrated at endline	None; baseline exploration index is the control	None; baseline conscientiousness and fluid intelligence are the controls
Adaptation and piloting	Stimuli replaced for the Indian context; piloted with 150 out-of-sample students	Item order randomized within the module	Responses grammar-corrected and translated before rubric scoring	Hard puzzles have 50 percent more empty cells; round 1 imposed hard for a random half

Notes: The table documents administration and scoring for the four skill-battery instruments of Section IV.C; the incentivized search task is documented separately in Appendix Table D.1. All instruments were administered on individual tablets within the endline battery, with module order randomized across students; both item-based tests also randomized item order. Cronbach's α is computed from the analysis sample's item responses (binary items, so α is the Kuder-Richardson coefficient; for the Sudoku task it is computed over the five round scores, and for idea generation over the scores of the three proposed ideas). Item-response thetas and all scores are standardized to the control group. The large-language-model scoring of the creativity tasks is detailed in Appendix Table D.3. Skill-to-instrument mapping and sources appear in Table 3.



FIGURE D.1. SAMPLE ITEM: READING THE MIND IN THE EYES TEST

Notes: Each item presents a photograph of the eye region of a face; the student selects, from four labeled options presented in English and Telugu, the emotion that best describes what the person is feeling or thinking. I adapted the 36-item RMET (Baron-Cohen et al. 2001) by replacing the original photographs with images appropriate to the Indian context. The adapted assessment was piloted with students outside the study sample. Item-level psychometric assessment using a two-parameter item-response model informed the selection of 20 items for baseline. The endline assessment used a shorter, 11-item form. Administration and psychometrics for the full battery appear in Table D.2.

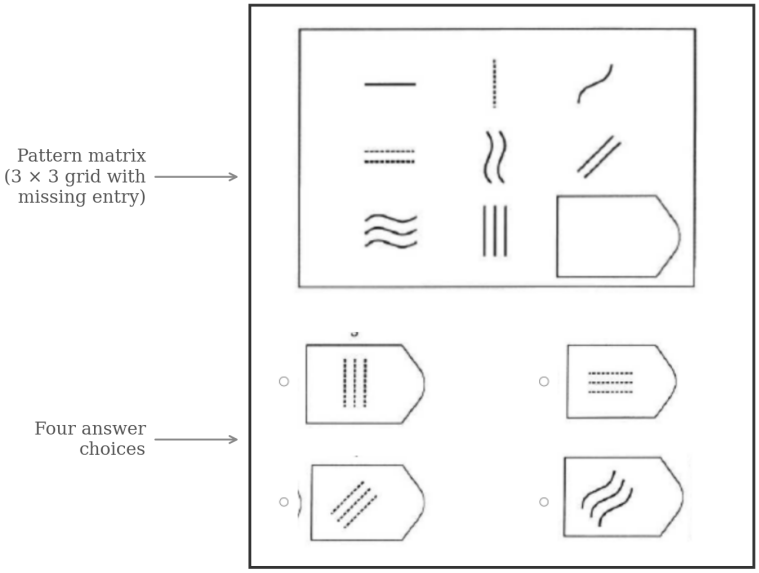


FIGURE D.2. SAMPLE ITEM: FLUID INTELLIGENCE (ICAR)

Notes: Each item presents a three-by-three grid of geometric patterns with one cell missing; the student completes the pattern from four or six options. The items are drawn from the International Cognitive Ability Resource (Condon and Revelle 2014). Following out-of-sample piloting, the baseline assessment used four response options rather than six to accommodate younger students returning after COVID-related school closures. Two years later, most endline items used the original six-option format, while selected items retained the four-option baseline format to preserve continuity across assessments. Administration and psychometrics for the full battery appear in Table D.2.

Panel A. Written expression



Write a creative story that connects the two images. Your story will receive a high score if it is original, imaginative, and well-written. You have five minutes to write this story and you should use 30-40 words.

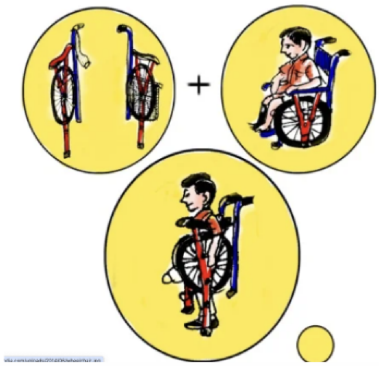
Panel B. Idea evaluation

You have to divide ₹1000 between two groups of students for their ideas below. Allot a total of ₹1000 but give more money to the idea which you think is more original and useful.

Torch with two bulbs



Wheelchair that can be folded to crutches



Tell us why you decided to distribute the money in this specific way. For example - why do you think one idea is better than the other?

Panel C. Idea generation



Try to imagine a "bicycle of the future". Think of 3 original improvements that can be made to a standard bicycle. The ideas should be different from each other and from the rest of your class. Explain how each idea works. Be creative!

FIGURE D.3. CREATIVITY TASKS: ORIGINAL STIMULI AND PROMPTS

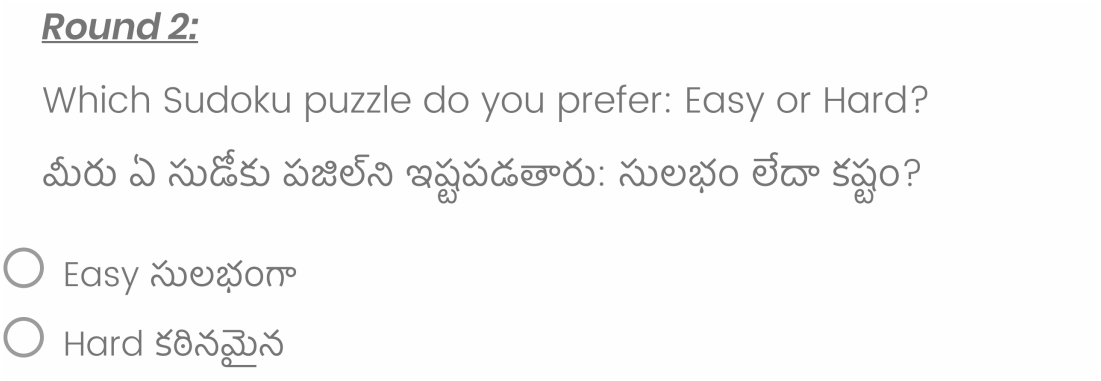
Notes: English prompts and original stimulus images from the February 2024 individual assessment, reformatted for display. Students received English and Telugu versions. Panels follow the order of the scoring rubrics in Table D.3; blank response fields are omitted.

TABLE D.3—PISA-ADAPTED CREATIVITY TASKS: SCORING RUBRICS AND EXAMPLE RESPONSES

Score	Rubric	Example Student Response	%
Panel A: Written Expression (Creative Story) (scored −1 to 5)			
-1	Non-interpretable or single word	*"Thdkfmgmgjyyigjgkg"*	21.8
0	Partial phrase or more than one word	*"The first image teaches the second image, then both teach together."*	8.2
1	One sentence with partial meaning	*"This is a picture with two images. There are arrows for some reasons."*	5.6
2	Multiple legible sentences without coherent idea	*"Tomorrow, I am going outside for shopping, and I want a toss. My friends have a toss, and my mom is very good."*	4.9
3	One meaningful sentence with image keywords (dice, globe, world, arrow, etc.)	*"Once upon a time on Earth, they were playing Ludo. One member won, and the other three members lost."*	25.4
4	Multiple sentences with partial/incomplete idea; no image keywords	*"Once upon a time, there was a girl named Niha and her friend. She was going to the office, and she would have a great..."*	14.7
5	Multiple sentences incorporating image keywords	*"In one universe, there was a planet named Earth. A big spaceship came and crashed into Earth, causing it to explode."*	19.5
Panel B: Idea Evaluation (Torch vs. Wheelchair) (scored 0 to 3)			
0	Non-interpretable or single word	*"Handicraft people are using bicycles."*	57.4
1	States preference or understanding without giving a reason	*"The first idea is better than the second one."*	32.0
2	Gives a reason but does not address cost, impact, ease, or user needs	*"Old cycles are used to prepare wheelchairs and for other purposes."*	5.8
3	Gives a reason about cost, help to users, impact, ease, or creative merit	*"With crutches, we can move very easily, but a torch can only be used at night."*	4.8
Panel C: Idea Generation (Bicycle Improvement) (scored 0 to 4)			
0	Non-interpretable or unrelated to bicycles	*"Gfkv"*	64.9
1	Mentions bicycle parts without suggesting improvement	*"I will improve the pedal."*	10.6
2	Very common features (lights, bell, brakes, basket, reflectors)	*"I will add a light to the bicycle."*	7.4
3	Common improvements beyond standard (electric motor, gears, suspension, speedometer)	*"I will add a comfortable seat."*	10.3
4	Unique or innovative (solar panels, distance sensor, auto-balancing, puncture-proof tires, energy-harvesting pedals, crash-alert notification, built-in first-aid kit)	*"I will add a distance sensor to the bicycle for crossing the road."*	6.9

Notes: The table reports task-specific rubrics and example responses for three PISA-adapted creativity tasks (OECD 2024). Percentages use 4,388 scored students in Panels A and B and 13,164 scored ideas in Panel C (three per student); non-interpretable responses enter at the bottom score level. Panel A scores a story about an image of dice, a globe, and an arrow from −1 to 5. Panel B scores the written justification of an allocation of ₹1,000 between two innovations from 0 to 3. Panel C scores each proposed bicycle improvement from 0 to 4; the student's outcome sums the three scores. Responses were grammar-corrected and translated to English using GPT-4o (temperature 0.3), then scored against the task-specific rubric (temperature 0). Examples are students' responses after grammar correction and illustrate the score levels; Section IV.C describes the scoring procedure. Administration details appear in Table D.2.

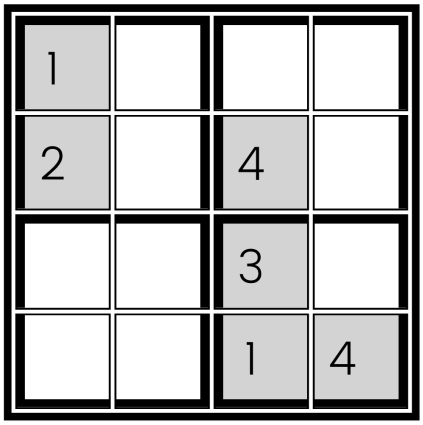
Panel A. Choice of puzzle difficulty



Please complete the following Sudoku puzzle

Fill the grid with numbers 1 to 4. Make sure each row, column, and 2x2 box has unique numbers. A number cannot appear more than once in a single row, column, or region.

Panel B. Easy puzzle



Panel C. Hard puzzle

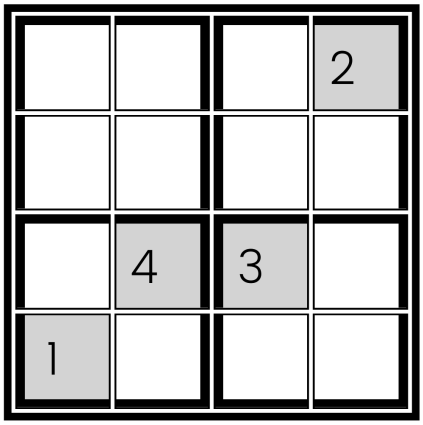


FIGURE D.4. SUDOKU TASK: DIFFICULTY CHOICE AND EXAMPLE PUZZLES

Notes: Panel A reproduces the round-2 difficulty-choice question from the February 2024 individual assessment. Panels B and C reproduce an easy and a hard puzzle from that round's item pools, with their common instructions above. Students had two minutes per puzzle. One round is selected at the end to determine the prize. Students receive a gel pen for solving a hard puzzle, a ballpoint pen for solving an easy puzzle, and no prize for an unsolved puzzle. Challenge-seeking is the share of rounds 2–5 in which the student chose the hard puzzle.

E. Academic Outcomes

TABLE E.1—COMPOSITION OF REPORTED WEEKLY TIME

Outcome	Treatment effect	SE	Sharpened q	N
Time on school periods (% of reported time)	-0.18	(0.87)	[0.504]	4,388
Time on homework (% of reported time)	-0.36	(0.45)	[0.269]	4,388
Time on sports (% of reported time)	0.79**	(0.38)	[0.068]	4,388
Preparation for the Innovation Challenge (% of reported time)	1.31***	(0.38)	[0.005]	4,388
Time on other activities (% of reported time)	-1.53**	(0.76)	[0.068]	4,388

Notes: The table reports treatment effects on the five activity shares of students' reported weekly time, in percentage points, for the 4,388 students who completed the endline assessment. Shares come from a five-activity weekly hours grid (school periods, homework and study, sports, preparation for the Innovation Challenge, and other activities), with each activity expressed as a share of the student's own total reported hours. Students could describe the other-activities category in a free-text field; entries in both arms include dance, singing, drawing, and reading. Every regression controls for randomization strata, school infrastructure, and individual covariates (age, socioeconomic index, computer skills), together with baseline mathematics and science scores and their missing-value indicators. Standard errors, in parentheses, are clustered at the school level, the level of treatment assignment; significance stars refer to the clustered p-value. Values in square brackets are Anderson sharpened false-discovery-rate q-values, computed within the five-share family. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

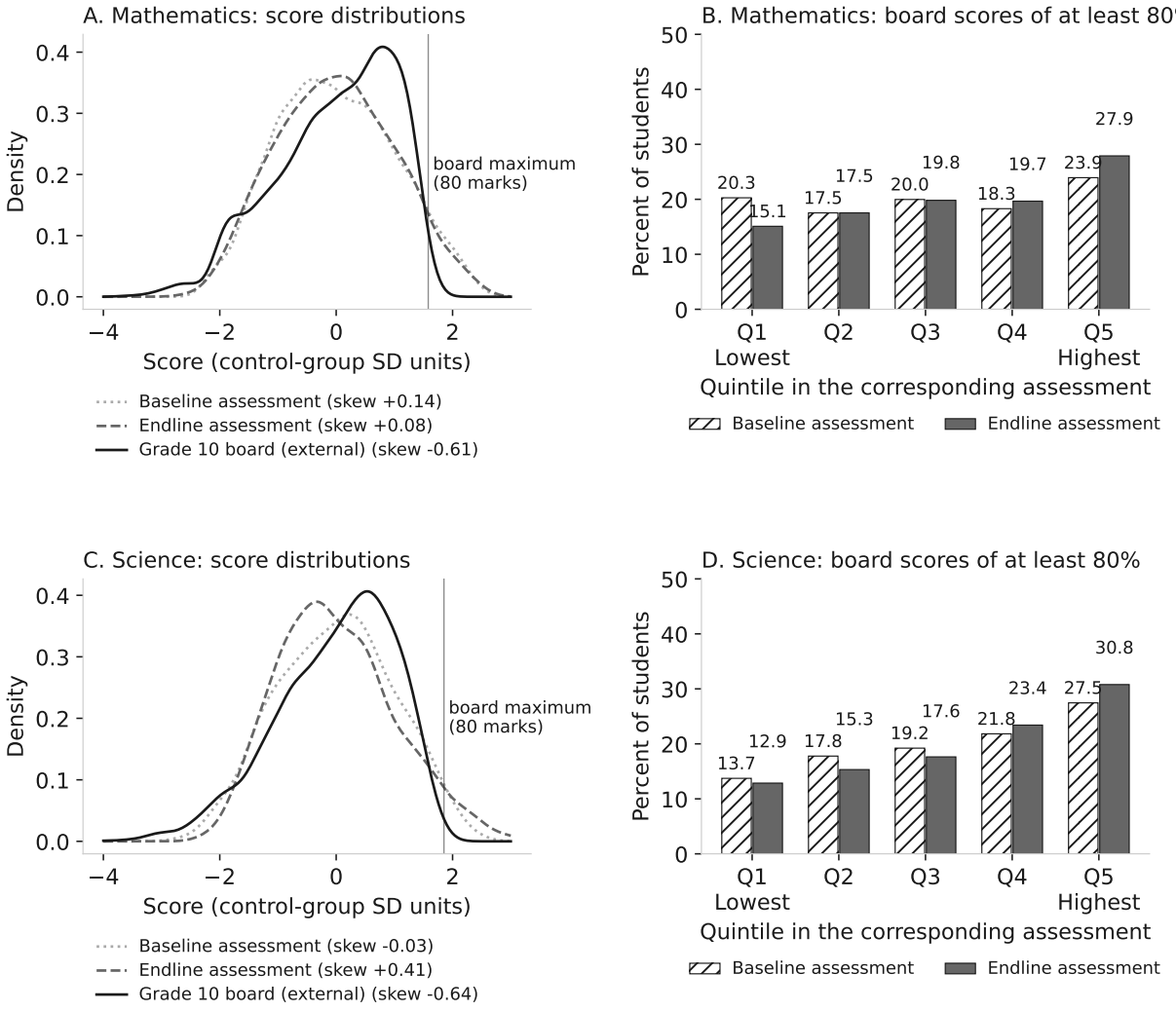


FIGURE E.1. CEILING COMPRESSION IN THE GRADE 10 BOARD EXAMINATIONS

Notes: Panels A and C compare the distributions of control students' mathematics and science scores at baseline (grade 8), endline (grade 9), and the grade 10 board examination. Each distribution uses all available observed scores, standardized within its own control-group sample. The vertical line marks the board examination maximum of 80 marks. Panels B and D focus on control students who scored at least 80% on the grade 10 board examination. Each bar shows the percentage of these students who belonged to a particular achievement quintile at baseline or endline. For example, among students scoring at least 80% in board mathematics, 20.3% had been in the lowest baseline quintile and 23.9% in the highest. Thus, high board marks were attained by students from across the earlier achievement distribution. Baseline and endline rankings are calculated separately within the same sample; each set of five bars sums to 100%. All scores are observed, without imputation.

TABLE E.2—MIDLINE ASSESSMENT, END OF YEAR 1

	Estimate	SE	*p*	*N*
Panel A. Mathematics				
Study cohort	−0.137	0.049	0.007	4,620
All students assessed at midline	−0.147	0.046	0.002	5,170
Panel B. Mathematics by subgroup, study cohort				
Social-welfare schools	−0.117	0.075	0.125	2,471
Tribal-welfare schools	−0.161	0.066	0.016	2,149
Difference	−0.045	0.102	0.662	
Boys	−0.161	0.068	0.020	2,057
Girls	−0.118	0.073	0.108	2,563
Difference	0.043	0.101	0.670	
Baseline mathematics below median	−0.138	0.054	0.012	1,785
Baseline mathematics above median	−0.119	0.079	0.136	1,781
Difference	0.018	0.082	0.824	
Panel C. Other outcomes, study cohort				
Science	0.000	0.048	0.998	4,620
Exploration index	0.145	0.052	0.006	4,620

Notes: The midline assessment was administered in February–March 2023, near the end of Year 1. Outcomes are in control-group standard deviations; standard errors are clustered by school. Panel A reports mathematics effects for students whose midline assessments are linked to grade 10 records and for all students assessed at midline. The broader sample also includes students without a linked grade 10 record, who may have left the school system before reaching grade 10. Comparing the estimates assesses whether the mathematics result changes when these students are included. Panels B and C use the study-cohort sample. Study-cohort regressions follow the Table 6 specification, using the corresponding baseline measure. The all-students regression controls for randomization strata, school infrastructure, midline age, household assets, computer use, and school-average baseline mathematics. Panel B reports subgroup effects from specifications that interact treatment with the subgroup indicator. Each "Difference" row subtracts the first-listed subgroup's treatment effect from the second; its *p*-value tests whether the two effects differ. The baseline mathematics comparison includes only the 3,566 students with observed baseline scores, split at their median. In Panel C, the exploration index combines the number of distinct configurations in the first four rounds and the number of distinct locations across the ten-round lemonade task.

F. Student Feedback at Midline

At the close of the Year 1 (midline) tablet assessment, treatment students answered one open-ended prompt, administered in English and Telugu: "Please write your thoughts about Inqui-Lab's Think and Make Program, through which you learned to work on two new ideas and products this year. What have you learned from this program?" Table F.1 summarizes the responses.

TABLE F.1—OPEN-ENDED STUDENT FEEDBACK AT MIDLINE: RESPONSE FUNNEL, VALENCE, AND THEMES

	N	Share
Panel A. Response funnel		
Students assessed at midline	2,319	
Wrote a non-empty comment	2,315	
Decipherable comment	2,231	
Substantive (at least one theme)	2,196	
Panel B. Overall valence (share of decipherable comments)		
Positive	2,032	91.1%
Neutral	192	8.6%
Negative	7	0.3%
Panel C. Content themes (share of substantive comments; multi-label)		
Enjoyment / affect	1,781	81.1%
Program-name mention (primed by the question)	731	33.3%
Problem-solving process language	386	17.6%
Generic learning claim	377	17.2%
Curriculum activity (sessions, making, models, teams)	313	14.3%
Assessment session (game, quiz, tablet, marks)	298	13.6%
Other themes	490	22.3%

Notes: The sample comprises students in treatment schools; the table is descriptive. Students left these comments in the final block of the Year 1 (midline) tablet assessment, in response to a single open-ended prompt (Appendix F). Panel A reports the response funnel: a comment is "decipherable" if it carries intelligible content (single letters and keyboard fragments are excluded) and "substantive" if it expresses at least one of the ten coded themes. Panel B shares are of decipherable comments. Panel C shares are of substantive comments; a comment may express more than one theme, so the shares sum to more than 100 percent, and *Other themes* collects the remaining, lower-frequency codes (a comment counts there if it carries at least one). Each comment was coded by a large language model, blind to treatment assignment, against a fixed codebook; the author read a random sample of the coded responses against the codebook. The codebook, the prompt, and the coded corpus are in the replication package.

The responses below, from students who received the curriculum, are lightly edited for readability and, where originally in Telugu, translated; the verbatim originals are in the replication package.

- *Problem-solving.* "I liked the Think and Make program and the Inqui-Lab Foundation because they are very useful for the development of the country. Children can take more initiative in solving problems. I am a student leader, and I gained a lot of experience in teaching problem-solving methods."

- *Making and community.* "I have learnt a lot from this Inqui-Lab. In this Think and Make program we made two prototypes. I learnt new things, and how to solve the problems around me and how to overcome them."

- *Change in school (translated from Telugu).* "I liked it very much because we worked as a group and had team discussions. Changes came to our school too: we started using steel bottles, and because of Inqui-Lab we stopped wasting resources. Thank you for this opportunity."

- *Transfer to daily life.* "I like this program because I think it is very innovative. I solved so many problems in the Think and Make program. This is really useful in our life: I learnt that if we have any problem in our surroundings, we can solve it."

- *Aspiration.* "I will find problems in our house, gurukulam, and school, and then find their solutions. Inqui-Lab's Think and Make program made me so happy. I want to become a scientist in my life. Thank you for this opportunity, sirs and madams."

G. Deviations from the Pre-Analysis Plan

The experiment was registered at the AEA RCT Registry as AEARCTR-0009696 in August 2022. The assessment instruments were uploaded in January 2024, immediately before endline data collection, but because of my oversight the registration's text was not revised to match them. The changes from the registration fall into four groups.

- **Registered instruments and outcomes omitted at endline.** The Innovators DNA survey was not administered, and the team outcome "community feedback" was not collected. The risk-taking task (Gneezy and Potters 1997) and the Big Five inventory were administered at baseline and midline but dropped from the endline battery. The risk task produced almost no variation at midline and the inventory is self-reported, so the assessment time was reallocated to performance-based tasks.

- **Instruments and data added after registration.** The three PISA-adapted creativity tasks and the timed puzzle task were not named in the original registration but were included in the uploaded instruments. Grade 10 board examination results were added as administrative data under a separate approval by the Stanford University IRB (protocol 79854, approved April 4, 2025, with waivers of consent and assent) and are reported as the academic boundary outcome (Table 6).

- **Instruments adapted or reinterpreted.** The registered exploration instrument (the "Lemonade Test" of Ederer and Manso) was adapted for tablet administration with three food stands, revised choice variables, and a class prize based on the highest profit achieved in any round. The adapted task preserves the registered construct of experimentation and tinkering, reported as the exploration outcome family in Table 5, Panel A. The puzzle task was intended to measure perseverance, but it provides no per-round feedback and therefore does not identify that construct. The paper instead reports the difficulty-choice margin (challenge seeking).

- **Analysis.** Heterogeneity in team outcomes is pre-specified in the trial registry, whereas the baseline-score splits are exploratory. Social-network measures and the registered team-composition analysis are reserved for a companion paper, since teams were randomized within schools in both arms, which constitutes an independent experiment.