# Can Bridging the Learning Gap Improve Test Scores on a Key National Exam? Evidence from a Remedial Education Program in Indian Secondary Schools

GAURAV CHIPLUNKAR
Darden School of Business, University of Virginia

DIVA DHAR
University of Oxford

RADHIKA NAGESH
Center for Global Development

## I. Introduction

Low- and middle-income countries across the world are facing a "learning crisis," wherein student learning outcomes have stagnated despite a significant increase in school enrollment (Glewwe and Muralidharan 2016; Singh 2024). In India, the focus of this study, this crisis is evident: only about 40% among a nationally representative sample of 14-to-18-year-olds could do a simple mathematical operation, while just over half (57.3%) could read sentences in English (Pratham 2023). These statistics reveal substantial gaps in foundational learning, raising concerns about the effectiveness of existing education systems in equipping students with essential skills.

Paradoxically, while students accumulate learning deficiencies as they progress through school, they still perform reasonably well on high-stakes national exams, which are supposed to assess advanced grade-level concepts.[1] For

This paper was previously circulated as "Not Too Little, but Too Late: Improving Post-Primary Learning Outcomes in India." We thank Prashant Bharadwaj, Jishnu Das, Alejandro Ganimian, Paul Glewwe, Clare Leaver, Anandi Mani, Isaac Mbiti, Karthik Muralidharan, Dipa Nag Chowdhury, Abhijeet Singh, as well as participants at the Research on Improving Systems of Education conference, Pacific Conference for Development Economics, Midwest International Economics Development Conference, and the Blavatnik School of Government and Centre for the Study of African Economies seminars for helpful comments and suggestions. We would also like to thank Avanti fellows, especially Akshay Saxena and Deepak Kamble. This work was supported by the MacArthur Foundation. All views expressed are our own and not of any of the institutions with which we are affiliated. Contact the corresponding author, Gaurav Chiplunkar, at chiplunkarg@darden.virginia.edu.

[1] Examples of national exams include the National Assessment of Educational Progress (NAEP) and the SAT in the United States, the General Certificate of Secondary Education (GCSE) in the United

Electronically published February 10, 2026

*Economic Development and Cultural Change*, volume 74, number 3, April 2026.

https://doi.org/10.1086/737565

example, half of the grade 9 students in our sample lacked basic numeracy skills; that is, they could not solve basic multiplication and division problems (see fig. 1 later in text). Yet the following year, the median student in this group achieved scores exceeding 50% on a standardized national exam that required them to solve calculus problems (see fig. 2 later in text). This disconnect raises concerns about the reliability of standardized national exams, which play a crucial role in shaping students' educational and employment opportunities (Das, Singh, and Chang 2022).

These observations motivate two key questions: (i) Can remedial education interventions implemented at the secondary level be cost-effective at improving student learning and, consequently, their performance on high-stakes national exams? (ii) Do national exams meaningfully assess student ability or do they obscure fundamental learning deficiencies? We examine these questions by evaluating a remedial education program implemented at the post-secondary-school level and benchmarking student performance across different assessment frameworks to better understand the relationship between remedial interventions, foundational learning, and standardized testing in high-stakes national exams.

We begin by reporting the results from our evaluation of an innovative after-school instructional program implemented with grade 9 students in 24 government schools in Chennai, India. Designed and delivered by Avanti (an education nonprofit), the program combined technology-aided instruction, peer learning, and counseling to address severe learning deficits in these grades (see sec. II for details). While extensive research has evaluated learning interventions at the primary school level in developing countries, less is known about remediation at the secondary level.[2] There is concern that remedial interventions potentially become less effective and more costly as students grow older, lag further behind grade-level requirements, and inevitably lose motivation. Our findings, however, suggest otherwise (see sec. IV.A). We show that remedial intervention at the secondary level led to large learning gains (around $0.8\sigma$–$0.9\sigma$) in basic literacy and numeracy, comparable with similar interventions targeting much younger students (Banerjee et al. 2016; Muralidharan, Singh, and Ganimian 2019). Despite our small sample and attrition, our results are robust across different specifications. Furthermore, at around 87 USD per standard deviation gain, the program was cost-effective as well, falling within the range (60–200 USD per standard deviation) of comparable interventions

Kingdom, the Kenya Certificate of Primary Education (KCPE) in Kenya, the All India Secondary School Examination (AISSE) in India, and so forth.

[2] For reviews on the primary level, see Kremer, Brannen, and Glennerster (2013), Evans and Popova (2016), and Glewwe and Muralidharan (2016).

(Banerjee et al. 2007; Kremer, Miguel, and Thornton 2009; Muralidharan, Singh, and Ganimian 2019). An additional concern with remedial programs is the possibility of diverting students' time away from grade-level material. The Avanti program required students to remain in school for one extra hour per day, 5 days a week, potentially affecting time allocated to schoolwork, tuition, and self-study. Nevertheless, we find no evidence of such trade-offs across multiple measures, including school absenteeism, participation in private tuition, parental support, and time spent on self-study (see sec. IV.C). Taken together, our results suggest that remedial programs at the secondary level can be both effective and economically viable.

Despite substantial gains in foundational learning, we find no effect of our intervention on test scores in a high-stakes, national exam—the All India Secondary School Examination (AISSE)—one year later.[3] Given that AISSE scores are crucial determinants of subsequent higher education admissions and employment opportunities, this finding is both important and concerning. However, it may not be entirely surprising—after all, how can students who struggle with basic arithmetic operations be expected to solve calculus problems within a year? Additionally, large-scale administrative exams in low-resource settings like India are often poorly implemented and face issues ranging from malpractice to grade inflation, which raises concerns about whether they can reliably measure true student ability (Singh 2024).

Our study provides us with a unique opportunity to leverage the Annual Status of Education Report (ASER) assessments to benchmark the effectiveness of AISSE exams in measuring student ability. We begin by comparing the cumulative distribution function (CDF) of the total AISSE score for students with minimal literacy (i.e., those who could only identify letters and words in the ASER reading test) with those for students who could read stories and paragraphs (see fig. 2*A* later). We conduct a similar comparison between students with minimal numeracy (i.e., those who could not perform basic arithmetic operations in the ASER math test) and students who could perform basic arithmetic operations in the ASER math test (see fig. 2*B* later). Both comparisons reveal a consistent set of insights: first, ASER and AISSE scores are positively correlated. Students with stronger foundational learning scored 20% higher on the AISSE as well. Second, even a median student with minimal literacy and numeracy skills (as defined above) scored over 50% on the AISSE, a

[3] As reported in table C3 (tables A1–C3 are available online), power calculations indicate that our experiment was statistically powered to detect effects of around 5–8 percentage points, which we considered reasonable given the empirical constraints. As reported in sec. V.A, our estimated effect ranged from 0.4 to 2 percentage points across subjects.

pattern that persists across subject-specific comparisons as well (see fig. A1 [available online]). This suggests administrative exams such as the AISSE are, at best, imperfect and noisy measures of student ability. The disconnect between learning and exam performance may stem from teaching practices that emphasize rote memorization or from poor implementation and grading inconsistencies, both of which align with existing critiques of the Indian educational system (Kingdon 2007; Muralidharan 2013; Pritchett 2013; Singh 2024).

Finally, we examine the heterogeneous effects of the intervention on academically weaker students, defined as those who could not do basic multiplication or division at baseline (see sec. V.C). The results yield both optimistic and cautionary insights. To begin, weaker students at baseline performed substantially worse at end line as well—across all metrics, including AISSE scores and ASER assessments (see table 5 later in text). On the optimistic side, the intervention did help narrow these gaps. It improved overall AISSE scores for weak students by 3.4 percentage points, closing around a third of the gap relative to their nonweak peers (table 5, col. 4). Subject-specific comparisons, on the other hand, provide a more nuanced insight—relatively larger gains in foundational math among weaker students in treated schools (col. 6) did not translate into higher AISSE math scores (col. 3). In contrast, AISSE language scores improved meaningfully (cols. 1 and 2), despite no differential gains in foundational reading among weaker students (col. 5).

Taken together, our analysis suggests that remedial programs can lead to better learning outcomes, even in advanced grades, and potentially among students with weaker foundational learning. However, the analysis highlights limited sensitivity of standardized exams to measure these conceptual gains, which are often affected by unrelated factors related to rote memorization, implementation, and so forth. The results therefore put into question whether improving learning skills alone can yield gains in these high-stakes assessments. Instead, they highlight the need to treat learning outcomes and test scores as distinct policy targets when evaluating the effectiveness of remedial education programs.

We contribute to four strands of the literature. First, we contribute to a growing body of research that has evaluated the effect of pedagogical interventions to improve learning outcomes and test scores. These include Teaching at the Right Level (Banerjee et al. 2007, 2016), computer-assisted learning (Linden, Banerjee, and Duflo 2003; Lai et al. 2015) including adaptive technology-aided instruction (Muralidharan, Singh, and Ganimian 2019), curriculum-based expert-led instructional videos (Beg et al. 2022), low-technology interventions such as SMS text messages and direct phone calls (Angrist et al.

2020), model schools (Kumar 2023), activity-based learning (de Barros et al. 2020), and performance-based incentives for teachers (Muralidharan and Sundararaman 2011). While this literature has primarily focused on improving outcomes in primary schooling, the focus of our study is to evaluate the effect of a similar remedial education program implemented at the post-primary level, where there is less evidence, especially in developing countries.

Second, there have been innovative variations in pedagogy to the standard remedial education programs studied in the literature. For example, Battaglia and Lebedinski (2022) study a program that matched students with older peers from same communities as mentors; Marinelli, Berlinski, and Busso (2024) study remedial tutorials that were offered during school hours in Colombia; and Saavedra, Näslund-Hadley, and Alfonso (2017) study remedial inquiry-based science education in after-school sessions in Peru. The model of remedial education that we study combines trained facilitators with technology aids, along with collaborative peer learning and teamwork among students. Moreover, it is specifically designed to be implemented in low-resource settings, similar to the models studied by other papers in this literature.

Third, our paper also complements evidence on "what works" in improving learning outcomes, such as provision of information on school quality (Andrabi, Das, and Khwaja 2015); student-level scholarships and incentives (Kremer, Miguel, and Thornton 2009; Blimpo 2014); improved access through free bicycles for girls (Muralidharan and Prakash 2017); curricular simplification and pacing (Pritchett and Beatty 2012; Mbiti and Rodriguez-Segura 2022); public rankings of schools (Cilliers, Mbiti, and Zeitlin 2021); and bundled interventions such as combining school grants with teacher incentives (Mbiti et al. 2019).

Finally, we contribute to the growing evidence on the disconnect between learning and test score performance on large-scale administrative exams (Kingdon 2007; Muralidharan 2013; Pritchett 2013). This literature is also not unique to India. Similar debates have emerged with respect to the effectiveness of SAT assessments in the United States (Cascio et al. 2024). While we do not prescribe an optimal design for such national exams, our paper contributes to the broader discourse by providing an alternative measure to benchmark their reliability in measuring student ability, similar to the paper by Singh (2024).

The rest of the paper is organized as follows: Section II describes the empirical context and the experimental design. Section III describes the data collection, sample, and our empirical specification. Section IV discusses the effect of our intervention on basic learning outcomes. Section V then explores whether these gains in basic learning translated into better scores on the AISSE,

highlighting a key disconnect in the assessment frameworks and potential reasons behind it. Finally, section VI offers a short conclusion.

## II. Empirical Context

### *A. Avanti's Program*

Avanti's blended learning approach for low-resource settings combines trained facilitators and technology aids and places an emphasis on peer learning and teamwork. There is also a small component involving parent and student counseling. Conducted after school for an hour for 5 days a week, Avanti's remedial classes run for 40 weeks in an academic year, led by trained facilitators rather than teachers. Facilitators are usually fresh graduates who are required to have a bachelor's degree, preferably (but not necessarily) in science, technology, engineering, and mathematics (STEM), and do not have teaching certification or qualifications. Instead, they are trained on facilitating classroom sessions, peer learning, and teamwork. They also rely on and provide prerecorded video lectures, presentations, worksheets, and other academic materials developed by Avanti for reading, math, and science, which are adapted to the context.[4] In the study schools, the program materials and discussions are mostly in Tamil (the local language), even though the school's official medium of instruction may be English or Tamil.

A typical facilitated class in the Avanti program includes one to two prerecorded, short 5-minute videos explaining key basic concepts in the grades 2–5 syllabus (such as carry-over addition, number line, solids and liquids, etc.). After each video, students are encouraged to gather in small groups to collaboratively solve worksheets and exercises with activities or multiple-choice questions related to the content of the video or lecture (e.g., sums, drawing on the number line, identifying solids, etc.). Students work together and help each other to problem-solve, with minimal intervention from facilitators, who only resolve or address questions when necessary. Finally, facilitators wrap up the session with lectures summarizing key concepts learned from the videos and in-class discussion. The program also facilitates occasional career counseling and guidance for students and parents. Table B1 provides a summary of the principal components of the Avanti program. The program cost approximately 45 USD per student every (school) year, or 5 USD per month for every student, which is

[4] Avanti's pedagogy incorporates features from Eric Mazur's peer instruction and collaborative learning pedagogy (Fagen, Crouch, and Mazur 2009; Schell, Lukoff, and Mazur 2013; Zhang, Ding, and Mazur 2017) and other blended learning programs. Avanti has adapted this methodology for low-resource settings (common to educational settings in developing countries and, in particular, India) for teaching reading, math, and science in government secondary schools in India by using basic technology and infrastructure available or installed in schools, such as computers (without internet).

cheaper than comparable secondary school programs in India such as MindSpark (15 USD per month) and those run by the government of Delhi (22 USD per month) (Muralidharan, Singh, and Ganimian 2019).

### *B. Experimental Design*

The study was conducted in a sample of 24 government schools under the management of the Chennai Municipal Corporation (CMC) in the city of Chennai, India. These 24 schools were selected from a total of 281 schools under the CMC, and their data were provided by the CMC Education Department. On the basis of the administrative data, the 24 schools were selected and deemed suitable according to the following eligibility criteria: (*a*) schools had operational grades 9 and 10 (65 schools); (*b*) there were a minimum of 45 students in grade 9;[5] (*c*) there were a maximum of 110 students;[6] and (*d*) there was no previous experience or engagement with Avanti (to avoid contamination).

The experimental design involved introducing the Avanti program in 12 randomly selected schools (henceforth, treatment schools) for the academic year 2016–17, leaving students from the remaining 12 schools to serve as a comparison group (control schools). The selection was done in an open paper lottery at the CMC under the observation of the research team to ensure that the selection was fair. Schools complied with the treatment selection, and the program was introduced for all grade 9 students attending the treatment schools. There was no further screening or targeting within treatment schools, and all students were required to attend the program. The control schools maintained a status quo, and no other program was introduced in those schools to our knowledge.

From each sample school, 45 students were randomly sampled on the basis of student registers at each school for a survey at the beginning of the academic year (henceforth, baseline). Out of 1,080 students listed from student registers, baseline surveys were completed with 991 students (92% of the listed sample). This was because some students could not be surveyed either because of school absence on the days the survey was taking place in their schools or because they had changed or dropped out of that school between enrollment and the baseline survey. Our final study sample therefore comprises these 991 students at baseline. During the year, some students dropped out or transferred to other schools resulting in a final (henceforth, end line) sample of 887 students 1 year later (89.5% of the baseline sample). In section IV.B, we find no evidence of

[5] Based on power calculations at 80% power and statistical significance at the 5% level, assuming a 10% sample attrition rate between the baseline and end line.

[6] For logistical reasons, Avanti could not handle schools with more than 110 students.

differential attrition between treatment and control groups, and the treatment effects are robust to constructing the corresponding Lee bounds (Lee 2009).

## III. Data and Empirical Specification

### *A. Sources of Data*

We use multiple datasets for our analysis that are a combination of primary surveys implemented by us and administrative data on student performance in a standardized national exam, the latter provided to us by our implementation partners. We discuss these data below.

*Student sociodemographic surveys*. As mentioned previously, we implemented two rounds of surveys with all students in our sample. Data from the baseline survey were collected from 991 students and subsequently from 887 students 1 year later during the end-line survey. We collected data on students' demographic and household information and aspirations. Given the program components focused on peer learning and teamwork, we also conducted a battery of tests to measure gender attitudes and noncognitive skills such as critical thinking, communication, goal setting, problem solving, grit, self-esteem, and teamwork (Duckworth and Quinn 2009; West et al. 2014). The survey tools were administered in the local language (Tamil).[7]

*Student learning outcomes (ASER tests)*. We measured learning outcomes of students at the start and end of the program using the standard ASER testing tool for reading and mathematics. The ASER instrument has been widely adopted in India (and other countries) to assess students' mastery over foundational skills at different stages of education. The ASER test uses a system to categorize students in five levels on the basis of the highest level reached and covers up to grade-2-level reading skills and up to grade-4-level mathematics ability (see table B2). The test was administered orally for each student by trained enumerators and lasted up to 10 minutes, following standard administration protocols laid out by ASER.[8] (Figure 1 later in text provides the distribution of the ASER reading and math skills for the students in our sample at baseline, as well as the appropriate grade-level requirement.)

*Guardian surveys*. For all students in the baseline survey, we also surveyed one of their guardians (either the mother or father) as well. They were asked questions on the support provided for the child's school activities in the household

[7] While we do not discuss the effect of our intervention on noncognitive skills (such as communication, problem solving, gender attitudes, etc.), we do measure them and find no effect. Results are available upon request.

[8] Information on the modalities of conducting the ASER test and tools is retrieved from https://asercentre.org/do-it-yourself-aser-diya/.

(see table B3), whether their child attended tuition classes after school, and the amount of time the child spent at home on self-study (such as doing homework, etc.). While we were able to survey a guardian for all students at baseline, because of various factors outside our control we were able to survey guardians for 856 out of 887 students at the end line (96.5%).

*Scores on a standardized national exam.* Finally, we also requested schools to provide the exam scores (across all subjects) for all students who appeared in a national standardized grade 10 examination in 2018. We use these data for multiple purposes: first, we match them to the students in our baseline data. This allows us to not only record their scores but also generate an indicator variable on whether a student appeared for the grade 10 exams or not. Second, because we have the scores across all students in these schools (irrespective of whether they were in our baseline sample), we also use them to examine the effect of our treatment more broadly. Finally, we combine these scores with the ASER to benchmark the usefulness of these exams in measuring student ability.

### B. Sample Description

We now describe our sample of students in panel A of table 1. Forty-one percent of students in our sample are female, and students are on average 14–15 years of age (which is appropriate for grade 9 students). About 88% come from socially disadvantaged backgrounds—Scheduled Castes (SC), Scheduled Tribes (ST), and Other Backward Castes (OBC)—and 76% are Hindu. From panel B of table 1, 20% of students have parents who are illiterate, and 65% have parents who have at most completed a high school degree. Only 10% of households earn a monthly income more than 10,000 INR (550 USD purchasing power parity). Panel C of table 1 describes the cognitive and noncognitive ability of students as well as the parental support they receive. As described before, we use the ASER tool to measure basic cognitive ability in reading and math (see table B2). The reading scores, for example, indicate that students are able to read words and paragraphs on average, but not stories. Similarly, for basic math, they could do simple operations of addition/subtraction, but not more complicated operations such as multiplication/division. These scores indicate that an average grade 9 student has only mastered skills usually taught in grades 2 and 3 and is lagging far behind in the grade-level skills needed. This finding resonates with other studies in the Indian context that document the wide learning gap at the post-primary level (Pratham 2023).

Columns 2 and 3 of table 1 report the averages separately for students in control and treatment schools. Column 4 reports the $p$-value from a $t$-test that examines whether these are statistically equal to each other. We do not see

**TABLE 1**
**SAMPLE CHARACTERISTICS**

| | Observations (1) | Control (2) | Treatment (3) | *p*-Value (4) |
|---|---|---|---|---|
| | | A. Student Characteristics | | |
| Female | 991 | .41 | .41 | .84 |
| Age | 991 | 14.85 | 14.92 | .22 |
| Hindu | 991 | .74 | .77 | .26 |
| SC/ST | 991 | .40 | .43 | .27 |
| OBC | 991 | .46 | .46 | .99 |
| | | B. Guardian Characteristics | | |
| Literacy: none | 991 | .21 | .19 | .59 |
| Literacy: high school | 991 | .65 | .64 | .82 |
| Income > 10,000 INR | 991 | .10 | .10 | .94 |
| | | C. Student Skills and Parent Support | | |
| ASER reading | 991 | 4.16 | 4.27 | .11 |
| ASER math | 991 | 4.49 | 4.52 | .52 |
| Parent support | 991 | 15.04 | 14.59 | .07 |
| Life skills | 991 | 22.99 | 23.01 | .87 |
| Student attitudes | 991 | 15.62 | 15.49 | .48 |

**Note.** Female takes the value 1 if the student is a female and 0 otherwise. SC/ST and OBC are indicator variables that take the value 1 if the student is from a Scheduled Caste/Scheduled Tribe or Other Backward Caste, respectively. Illiterate and high school take the value 1 if the student's guardian is illiterate or has completed at most high school education, respectively, and 0 otherwise. Income > 10,000 INR takes the value 1 if the guardian's income is more than 10,000 INR and 0 otherwise. See app. B for measurement of ASER and parent support variables. Column 4 reports the *p*-value of a *t*-test of the difference of means between treatment and control, i.e., cols. 2 and 3.

significant differences between students in treatment and control schools. The differences are small and statistically insignificant at conventional levels.

### C. *Empirical Specification*

To examine the effect of our intervention, we estimate the following regression specification for an individual $i$ in school $s$:

$$Y_{iE} = \alpha + \beta T_s + \gamma_1 Y_{iB} + \delta_1 X_i + \delta_2 X_s + \varepsilon_i, \quad (1)$$

where $Y_{iE}$ and $Y_{iB}$ are the outcome variables of interest for an individual $i$ measured at end line and baseline, respectively; $T_s$ is a binary indicator that takes the value 1 (0) for a treatment (control) school; and $X_i$ and $X_s$ are time-invariant characteristics of the individual and school, respectively. We use gender, age, religion, and a dummy for caste (SC/ST, OBC, and others) for $X_i$ and number of boys and girls and the language of instruction in the school for $X_s$. We estimate the above specification both without and with individual and school controls and report them in panels A and B, respectively, of tables 2, 3, and 4. Finally, because the randomization was done with 24 schools, we wild-bootstrap

cluster our standard errors at the school level, as suggested by Cameron, Gelbach, and Miller (2008) for statistical inference with small clusters. This is reported in tables 2–5 as the "*p*-val (OLS)" value. Furthermore, we also report the *p*-value from a two-sided randomization inference test, denoted as "*p*-val (RI)" in all tables. This test, originally proposed by Fisher (1935) and developed by Heß (2017) and Young (2019), allows for statistical inference by comparing the realized treatment effect with 1,500 placebo assignments. This procedure therefore has the advantage of providing inference with correct size, regardless of the sample and cluster size.

## IV. Effect on Basic Learning Outcomes

### *A. Effect on ASER Scores*

We start by examining the effect of the treatment on basic learning outcomes, as measured by the ASER reading and math scores. We examine the effect of our intervention in three ways. First, we look at the differences in ASER scores of students in the treatment and control schools at end line (cols. 1 and 2 of table 2). As mentioned previously, panel A reports the results without any individual and school characteristics, while panel B controls for them. Both panels control for a student's corresponding ASER score at baseline.

We make two observations before discussing the results. First, baseline and end-line scores are strongly positively correlated, and second, controlling for individual and school characteristics does not substantially alter the magnitude or statistical inference of the effect of the intervention. Turning to the effect of our intervention, from panel B of table 2 (our preferred specification), after we control for baseline ASER scores, individual and school characteristics, students in treatment schools have $0.92\sigma$ and $0.80\sigma$ higher ASER reading and math scores, respectively, compared with those of students in control schools at end line. As a benchmark, this effect is comparable to that of similar programs evaluated in the Indian context by using ASER, such as Teaching at the Right Level or the Balsakhi programs (Banerjee et al. 2007, 2016) or using other tests such as the Mindspark program (Muralidharan, Singh, and Ganimian 2019). Because the ASER scores are discrete and ordinal, we also use an alternate specification and report results from a rank-ordered logistic regression in columns 3 and 4 of table 2. The results and their statistical inference are qualitatively similar for both outcome variables.

We then examine the distributional effect of the intervention across different levels of cognitive ability. Specifically, in table A1, we report the difference in the end-line cognitive level by comparing treatment and control students with the same cognitive level at baseline. For example, in column 1 of table A1, we restrict the sample of students (in both treatment and control schools) to only

**TABLE 2**
**EFFECT ON COGNITIVE ABILITY**

| | Standardized ASER Scores | | ASER Score | |
|---|---|---|---|---|
| | Read (1) | Math (2) | Read (3) | Math (4) |
| | A. Without Individual and School Controls | | | |
| Treatment | .93*** | .80*** | 2.03*** | 1.96*** |
| | (.13) | (.14) | (.29) | (.32) |
| Baseline | .46*** | .54*** | .89*** | 1.45*** |
| | (.05) | (.05) | (.10) | (.13) |
| *p*-val (OLS) | .00 | .00 | .00 | .00 |
| *p*-val (RI) | .00 | .00 | .00 | .00 |
| $R^2$ | .36 | .30 | | |
| | B. With Individual and School Controls | | | |
| Treatment | .92*** | .81*** | 2.04*** | 2.05*** |
| | (.16) | (.16) | (.36) | (.39) |
| Baseline | .44*** | .52*** | .88*** | 1.39*** |
| | (.05) | (.06) | (.11) | (.14) |
| *p*-val (OLS) | .00 | .00 | .00 | .00 |
| *p*-val (RI) | .00 | .00 | .00 | .00 |
| $R^2$ | .37 | .32 | | |
| Observations | 887 | 887 | 887 | 887 |
| Regression type | OLS | OLS | Ordered logit | Ordered logit |

**Note.** Outcome variables in cols. 1 and 2 are the ASER scores for reading and math that have been standardized to have mean 0 and standard deviation 1 for the control group. Individual controls include gender, age, religion, and caste. School controls include language of instruction and the number of boys and girls in the school. "Baseline" captures the students' responses to the survey question at baseline. Wild-bootstrapped standard errors are clustered at the school level and reported in parentheses. *p*-val (OLS) reports the *p*-value for the treatment coefficient estimated by the wild-bootstrapped clustered standard errors, while *p*-val (RI) reports the *p*-value using the randomized inference method. OLS = ordinary least squares; RI = randomized inference.
*** $p < .01$.

those who could read at least a word at baseline. We find that these students in treatment schools have a 0.43 higher ASER reading score at end line compared with that of their control school counterparts. Similarly in column 2, treatment school students who could read a paragraph at baseline have a 0.78 higher ASER score at end line compared with that of their control school counterparts who could also read only a paragraph at baseline. Finally, in column 3, we see that among students who could read a story at baseline, those in treatment schools (compared with control) have about 0.76 higher reading score at end line. Similarly, turning to math scores in columns 4–6, we find that students in the treatment schools have a higher math score at end line compared with that of their control school counterparts who start at the same baseline math level. Put together, the above analysis suggests that the Avanti program robustly improves students' basic reading and math skills across all levels of competencies.

### B. Attrition

As mentioned previously, we were able to survey 887 students from our baseline sample of 991 students (an attrition of 10.5%). This attrition was caused both by students dropping out of school and by transfers to other schools (e.g., due to family migration). In appendix C, we test for differential attrition between control and treatment groups and check whether this affects our estimates. Specifically, we do so in two ways: first, we examine whether student and parent characteristics, cognitive and noncognitive ability, and parental support are differentially correlated with dropout and transfer (table C1). We find no evidence of differential attrition. Second, to account for potential endogenous attrition from the sample, we estimate Lee bounds on the treatment effects (Lee 2009). We report the 90th, 95th, and 99th percentiles of the lower and upper bounds for our reading and math scores respectively. These estimates are consistent with our main analysis. For example, from panel A of table C2, the attrition-adjusted lower bound on the effect of our intervention lies between $0.59\sigma$ and $0.75\sigma$ for reading and $0.26\sigma$ and $0.37\sigma$ for math. This increases our confidence in the robustness of our results after taking into account attrition of students from our sample.

### C. Did the Intervention Crowd Out Time from Grade-Level Studies?

The Avanti program required students to remain in school for one additional hour per day, 5 days a week. A key concern with implementing such a remedial education program is that while it may enhance foundational learning (as documented above), it could also divert students' time and effort away from grade-level material. This concern is particularly salient in advanced grades, such as in our study setting, where students are expected to prepare for a high-stakes national exam the following year.

Using data from the guardian survey (described in sec. III), we examine the effect of our intervention across a wide range of measures of students' time and effort outside the school, ranging from absenteeism in school, parental support in studying to participation in private tuition classes, and time spent on self-study after school. As reported in table 3, we find no effect of our intervention on any of these activities. The estimated coefficients are small in magnitude and statistically insignificant at conventional levels. If anything, the estimated coefficients in some cases go in the opposite direction of what would be needed to explain away our results through these channels.

For instance, we find no effect of our intervention on absenteeism in school, as measured by the number of days that a student was absent in the previous week (table 3, col. 1). The program had no effect on parental support for learning

**TABLE 3**
**EFFECT ON STUDENTS' EFFORT ON STUDIES, PARENTAL SUPPORT**

| | | | Tuition | | Self-Study | |
|---|---|---|---|---|---|---|
| | Absenteeism (1) | Parent Support (2) | At Least 1 Day (3) | Days/Week (4) | At Least 1 Day (5) | Days/Week (6) |
| | A. Without Individual and School Controls | | | | | |
| Treatment | −.07 | .18 | −.02 | −.11 | −.08 | −.74 |
| | (.17) | (.38) | (.03) | (.20) | (.08) | (.88) |
| Baseline | .18*** | .28*** | .46*** | .44*** | .01 | .12*** |
| | (.04) | (.04) | (.04) | (.04) | (.04) | (.04) |
| *p*-val (OLS) | .66 | .63 | .57 | .58 | .33 | .40 |
| *p*-val (RI) | .68 | .63 | .56 | .56 | .35 | .42 |
| $R^2$ | .03 | .07 | .21 | .19 | .01 | .02 |
| | B. With Individual and School Controls | | | | | |
| Treatment | −.10 | .22 | −.01 | −.08 | −.08 | −.74 |
| | (.19) | (.44) | (.04) | (.23) | (.08) | (.92) |
| Baseline | .16*** | .29*** | .46*** | .43*** | .01 | .10** |
| | (.04) | (.04) | (.04) | (.04) | (.04) | (.04) |
| *p*-val (OLS) | .62 | .62 | .71 | .72 | .35 | .42 |
| *p*-val (RI) | .61 | .58 | .68 | .66 | .36 | .42 |
| $R^2$ | .05 | .10 | .21 | .20 | .05 | .07 |
| Observations | 856 | 856 | 856 | 856 | 856 | 856 |

**Note.** "Absenteeism" is the number of days in the past week that a student was absent from school. "Parent support" is a composite index described in app. B. Tuition in col. 3 takes the value 1 if a student takes private tuition classes and in col. 4 is the number of days/week that the student attends tuition classes. Self-study in col. 5 takes the value 1 if a student spent at least 1 day in the past week self-studying and 0 otherwise and in col. 6 is the number of days/week that a student self-studies. Individual controls include gender, age, religion, caste, and the baseline value of the outcome variable. School controls include language of instruction and total number of boys and girls. Wild-bootstrapped standard errors are clustered at the school level and reported in parentheses. p-val (OLS) reports the *p*-value for the treatment coefficient estimated by the wild-bootstrapped clustered standard errors, while *p*-val (RI) reports the *p*-value using the randomized inference method.
** $p < .05$
*** $p < .01$.

either (col. 2).[9] If anything, the estimated coefficients indicate a small reduction in absenteeism (by around 6%) and a marginal increase in parental support (by around 1%), though neither is statistically significant at conventional levels. Similarly, in columns 3–6, we find no effect of the intervention on either the probability of attending private tuition classes or studying at home (cols. 3 and 5, respectively) or the amount of time spent on either activity (cols. 4 and 6, respectively).

Taken together, we conclude that the intervention was able to increase students' foundational learning without crowding out their effort and time toward

[9] The questions used to construct the index of parental support are described in app. B. It aggregates responses on a Likert scale from 1 to 5 across four indicators of how often parents help their child with schoolwork, homework, attending events, and talking about school activities. As reported in table A2, the intervention has no effect on any of these individual indicators either.

studying grade-level material. In the next section, we then turn our attention to examining whether these gains in foundational learning also helped improve test scores on a high-stakes national exam a year later.

## V. Did Gains in Basic Learning Translate into Better Scores on a Standardized National Exam?

It is mandatory for all students in India completing grade 10 to take a standardized national exam: the AISSE. Students are tested in five subjects, namely: math, English, science, social sciences, and the local language (Tamil, in our case). The AISSE is an important milestone in a student's education because these scores are used for future admissions in higher education and for early career job applications. One of the reasons that Avanti's program was delivered in grade 9, at the behest of the government, was to help students perform better in this grade 10 national exam.

### *A. Effect on AISSE Scores*

As noted earlier, we obtain data on AISSE scores for all five subjects for all students in our 24 schools and begin by matching them to the students in our study sample. The results are reported in table 4. Consistent with previous analysis, we

**TABLE 4**
**EFFECT ON STANDARDIZED NATIONAL EXAM SCORES**

| | Any Exam? (1) | Tamil (2) | English (3) | Math (4) | Science (5) | Social Science (6) | Total (7) |
|---|---|---|---|---|---|---|---|
| | A. Without Individual and School Controls | | | | | | |
| Treatment | −.00 | .54 | 1.37 | 1.92 | −.65 | −.97 | .41 |
| | (.05) | (2.99) | (2.40) | (2.33) | (2.50) | (3.07) | (2.03) |
| *p*-val (OLS) | .99 | .86 | .57 | .41 | .80 | .75 | .84 |
| *p*-val (RI) | .99 | .86 | .59 | .42 | .80 | .75 | .86 |
| $R^2$ | .00 | .00 | .00 | .01 | .00 | .00 | .00 |
| | B. With Individual and School Controls | | | | | | |
| Treatment | .00 | −.82 | .65 | 1.85 | −1.32 | −2.00 | −.35 |
| | (.04) | (2.13) | (2.31) | (2.10) | (2.37) | (2.45) | (1.57) |
| *p*-val (OLS) | .99 | .70 | .78 | .38 | .58 | .41 | .82 |
| *p*-val (RI) | 1.00 | .71 | .79 | .41 | .59 | .40 | .84 |
| Control mean | .82 | 64.56 | 51.22 | 45.68 | 62.35 | 59.79 | 56.77 |
| $R^2$ | .14 | .28 | .20 | .13 | .17 | .21 | .25 |
| Observations | 991 | 814 | 814 | 814 | 814 | 814 | 814 |

**Note.** All outcome variables are the percentage marks scored by a student in his or her AISSE exams. Panel A does not include any controls, while panel B includes individual and school controls. Individual controls include gender, age, religion, and caste. School controls include language of instruction, number of boys and girls, and the baseline ASER scores for reading and math. Wild-bootstrapped standard errors are clustered at the school level and reported in parentheses. *p*-val (OLS) reports the *p*-value for the treatment coefficient estimated by the wild-bootstrapped clustered standard errors, while *p*-val (RI) reports the *p*-value using the randomized inference method.

report the results without any controls in panel A and with students and school controls, along with baseline reading and math ASER scores, in panel B. In column 1 of table 4, we construct a variable that takes the value 1 if a student in our sample appeared for the AISSE and 0 otherwise. Eighty-two percent of students in our sample appeared for the AISSE, and the intervention had no effect on increasing this probability. In columns 2–7, we then examine the effect of our intervention on AISSE scores. Reassuringly, we see that baseline ASER scores are strongly positively correlated with AISSE scores. However, the treatment does not have any significant effect on these exam scores. The estimated magnitudes are very small (±0.4 to 2 percentage points) across all subjects and nowhere statistically significant at conventional levels.[10]

In table A3, we take advantage of the fact that we have the AISSE scores across all students in our sample schools, not just the students that we sampled in our surveys. This allows us to examine the effect of our intervention across the school more broadly. However, gender and caste are the only two student demographic variables available in these data that we can control for in our regression specification. As reported in table A3, we find no effect of our intervention on AISSE scores. The estimated coefficients are very small (±1 to 2 percentage points), statistically insignificant at conventional levels, and comparable to our study sample (in table 4).

### *B. ASER Assessments and Effectiveness of AISSE in Measuring Student Ability*

The above results, while disappointing, are perhaps not that surprising. After all, how can students who struggle with basic arithmetic operations be expected to solve grade-level calculus problems? In addition, large-scale administrative exams in resource-constrained settings like India are often poorly implemented, raising concerns about whether they accurately measure true student ability (Singh 2024). Distinguishing between these two explanations requires an independent benchmark to assess the reliability of administrative exam results—a challenge that our study is uniquely positioned to address. We leverage two distinct measures of student ability—AISSE national exam scores and ASER test scores in reading and mathematics—to learn more about the effectiveness of standardized national exams in assessing grade-level proficiency.

As shown in figure 1, half of the students in our sample entering grade 10 lacked basic numeracy or literacy skills; that is, they could not either perform basic multiplication and division or read simple paragraphs or stories according

[10] As reported in table C3, power calculations indicate that our experiment was statistically powered to detect effects of around 5–8 percentage points, which we considered reasonable given the empirical constraints.

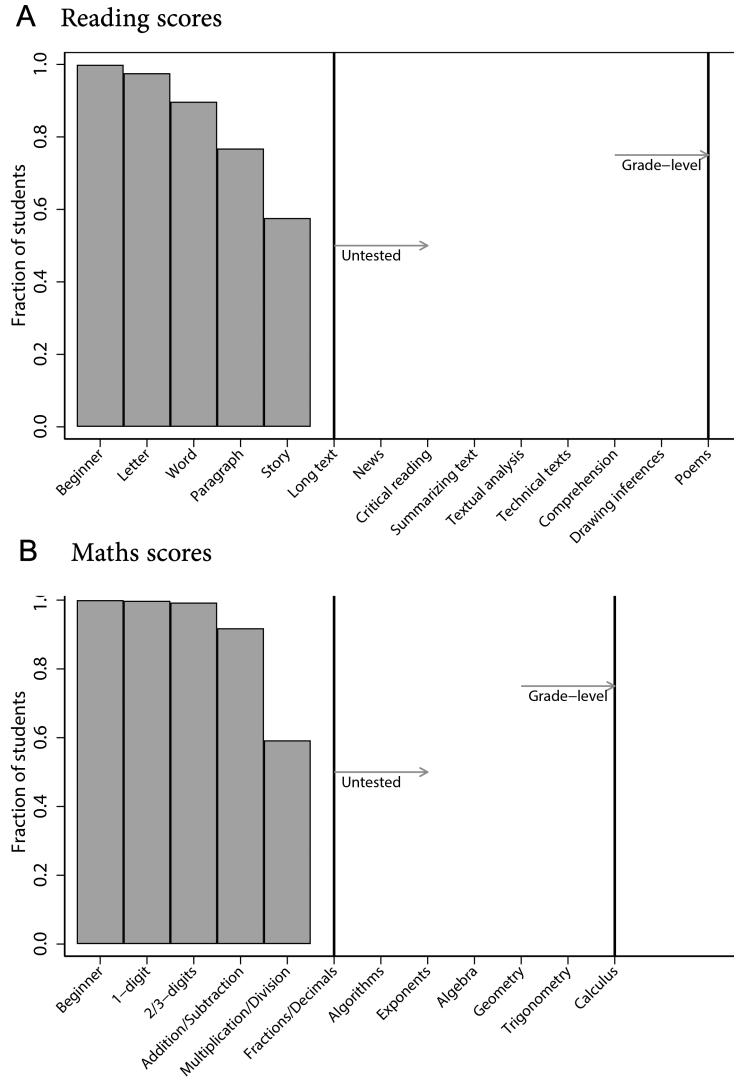


**Figure 1.** Reading and math levels at baseline. *A*, Reading scores. *B*, Math scores. Shown is the cumulative fraction of students across various levels of reading and math as measured by the ASER test. Details on the ASER survey are provided in appendix B.

to their ASER assessments. This finding is not unique to our sample and is consistent with the learning levels among a nationally representative sample of 14- to-18-year-old students (Pratham 2023). Figure 2 shows how these learning deficits translate into AISSE performance less than a year later. Specifically, figure 2*A* plots the cumulative distribution function (CDF) of the total AISSE score, comparing students with minimal literacy (i.e., those who could only

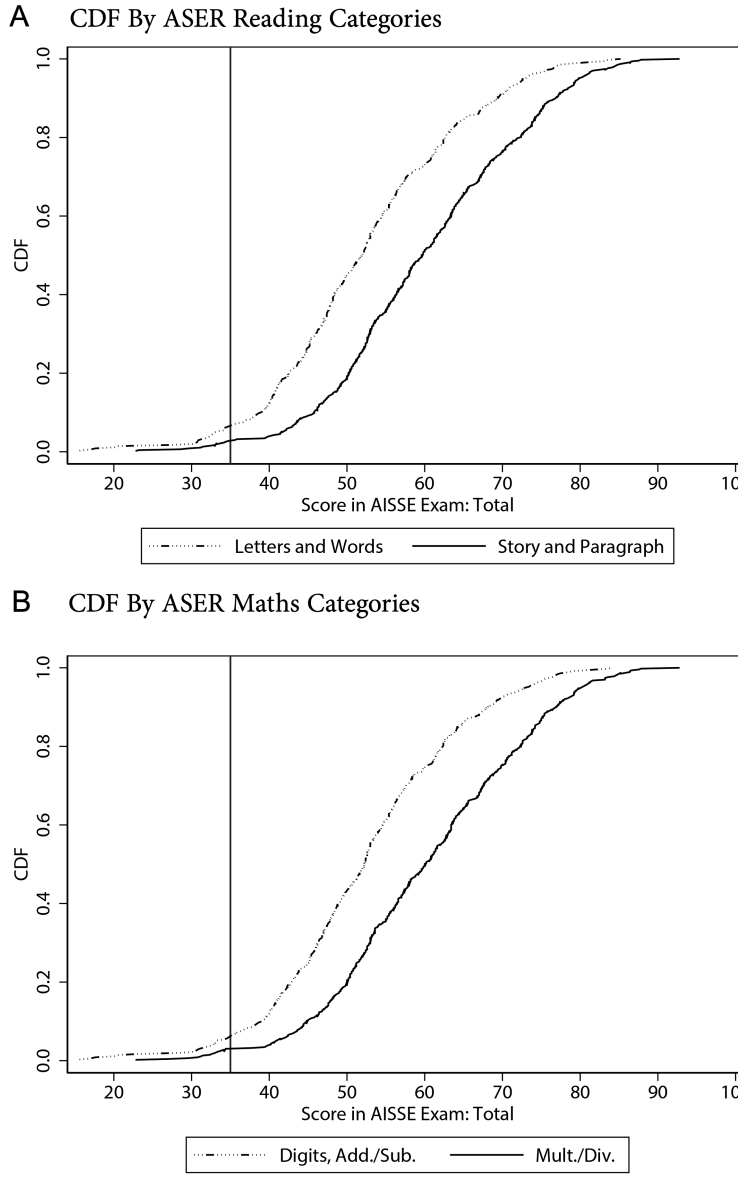


**Figure 2.** Cumulative distribution function (CDF) of AISSE scores by ASER reading and math ability. *A*, CDF by ASER reading categories. *B*, CDF by ASER math categories. Shown is the CDF of the total score in the AISSE by students. *A*, The CDF is shown separately for students who could read letters and words and those who could read stories and paragraphs as measured in the end-line ASER reading test. *B*, The CDF is shown separately for those could at most identify digits or complete addition/subtraction as opposed to those who could complete multiplication/division as well, as measured by the end-line ASER math test. Details on the ASER survey are provided in appendix B. The vertical line shows the minimum marks required to pass the exam.

identify letters and words in the ASER reading test) with those who could read stories and paragraphs. Figure 2*B* presents a similar CDF analysis for students who had minimal numeracy (i.e., those who could not perform basic arithmetic operations in the ASER math test) compared with students who could perform basic arithmetic operations in the ASER math test.

Our findings reveal two key insights. First, ASER and AISSE scores are positively correlated: students with stronger foundational learning scored 20% higher on the AISSE as well; and second, even students with minimal literacy and numeracy skills (as defined above) scored over 50% on the AISSE. This pattern persists even across subject-specific comparisons as well, such as ASER reading and AISSE language scores (fig. A1*a*) and ASER and AISSE math scores (fig. A1*b*).[11]

These patterns suggest that large-scale, administrative exams like the AISSE are, at best, imperfect and noisy signals of underlying student ability. Several factors may contribute to this disconnect between the AISSE scores and ASER learning outcomes. One possible explanation is that students are encouraged to rely on rote memorization and recall of textbook questions, as opposed to conceptual understanding, and that schools are also preparing students for the AISSE by "teaching to the test." This is in line with existing critiques of the Indian educational assessment system (Kingdon 2007; Muralidharan 2013; Pritchett 2013). Another explanation could be grade-inflation or exam-related malpractice, widely prevalent in low-capacity settings like India (Singh 2024).[12] More broadly, our findings suggest that when scaling up educational interventions based on learning outcomes and skills, the measurement constraints can themselves become a critical challenge given the difficulties in using administrative data to gauge or measure true learning. Without accurate assessments, policy makers may risk misallocating resources or misinterpreting the effect of interventions.

### C. *Heterogeneous Effect on Academically Weaker Students*

We now turn to examining whether the intervention had a differential effect on academically weaker students in terms of their AISSE performance. The expected effect is theoretically ambiguous: while weaker students stand to gain the most from remedial interventions, they may also face higher marginal costs of deviating from grade-level curriculum, making it harder to keep up with coursework.

[11] Notably, fig. A1*b* shows a distinct clustering of AISSE math scores around the passing threshold. We do not observe similar bunching in other subjects. While there could be many potential explanations for this, ranging from malpractice to grade inflation, our data do not allow us to conclusively determine the underlying cause of this pattern.

[12] This could explain bunching around the passing threshold as observed in fig. A1*b*.

To examine this, we define a binary variable, "weak," which takes the value 1 if a student lacked basic numeracy skills (i.e., was unable to perform a simple multiplication or division task at baseline). By this criterion, approximately 40% of students in the study are classified as weak.[13] Figure 3 plots the empirical CDFs of the total AISSE score disaggregated by weak and nonweak students in control and treatment schools. As expected, we find a significant gap in AISSE scores between weak and nonweak students (consistent with our discussion above). However, there is also a distinct rightward shift in the CDF for weak students in treated schools relative to their counterparts in control schools. In contrast, no such noticeable difference arises among nonweak students. We formally examine this insight by estimating a regression specification, similar to equation (1), for a student $i$ in school $s$:

$$Y_i = \alpha + \beta_T T_s + \beta_W \text{Weak}_i + \beta T_s \times \text{Weak}_i + \delta_1 X_i + \delta_2 X_s + \varepsilon_i, \quad (2)$$

where $\beta_W$ estimates the average difference in the outcome variable ($Y_i$) between weak and nonweak students in control schools, while $\beta$ captures the differential effect of the intervention for weak relative to nonweak students in treated relative to control schools.

The results, reported in table 5, offer mixed but instructive insights. First, they confirm that academically weaker students in control schools perform substantially worse across all outcomes, including AISSE scores and foundational learning assessments at end line. For instance, weaker students score 20% lower in English and Tamil, 12% lower in math, and 17% lower overall on the AISSE (cols. 1–4). They also lag significantly in foundational skills, with end-line ASER reading and math scores that are $0.5\sigma$ and $1.1\sigma$ lower, respectively (cols. 5 and 6).

Second, and more optimistically, we find that the program helped narrow these gaps. Column 4 shows that the intervention improved overall AISSE scores of weaker students by 3.4 percentage points, closing 36% of the preexisting gap relative to nonweak students.[14]

[13] There is a strong correlation between ASER reading and math scores among weak and nonweak students. For example, 61% of numerically weak students were also weak in literacy, meaning they could not read a story, compared with 70% of numerically nonweak students who could. More broadly, students classified as numerically weak had a $0.67\sigma$ lower ASER reading score at baseline relative to their numerically nonweak counterparts.

[14] We calculate this as the ratio of $\hat{\beta}/\hat{\beta}_W$ in eq. (2) or (Treat × Weak)/Weak in table 5.

Table A4 provides suggestive evidence of some behavioral mechanisms. The number of absent days per week decreased by 28% (col. 1), and parental support increased by 6.3% (col. 2) for weaker students relative to nonweak students. The intervention had no differential effects on time spent on tuition or self-tudy, though the point estimate on the latter is quantitatively large (15%) but statistically insignificant at conventional levels.

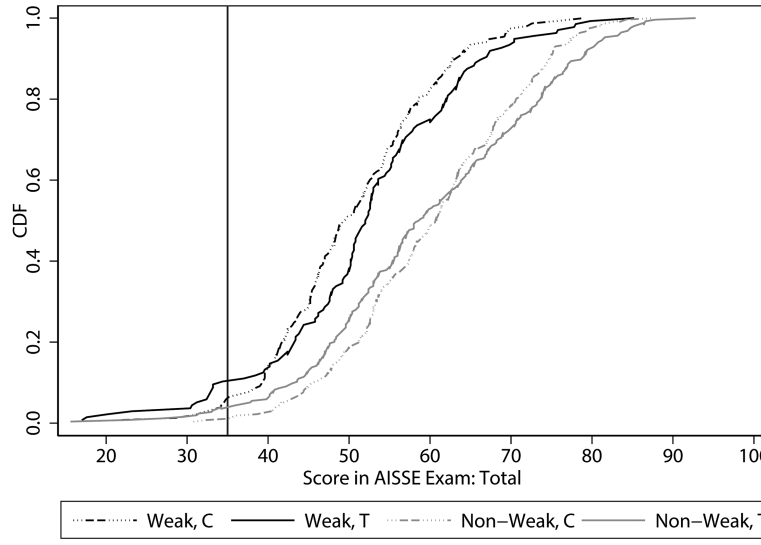


**Figure 3.** Cumulative distribution function of AISSE performance for academically weak and nonweak students. Plotted is the CDF of the total AISSE scores separately by academically weak and nonweak students in control and treatment schools. The vertical line shows the minimum marks required to pass the exam.

Third, subject-specific results provide a more mixed insight. Despite large improvements in foundational math among weaker students ($0.55\sigma$ in ASER; col. 6 of table 5), we detect no corresponding gains in AISSE math scores (col. 3). Conversely, we find meaningful reductions of 4–6 percentage points (nearly 40% of the gap) in AISSE language scores (cols. 1 and 2), even though weaker students did not exhibit differential improvements in foundational ASER reading (col. 5). This asymmetry echoes the broader disconnect we discuss in section V.B, where students with minimal ASER proficiency still scored reasonably well on the AISSE and where foundational gains did not always predict exam outcomes.

Taken together, our results reflect the limited sensitivity of standardized exams to measure real conceptual gains as opposed to the effect of unrelated factors (such as rote preparation, test-taking strategies, grading practices, etc.). They question the assumption that improving basic skills alone will yield parallel gains in high-stakes assessments and instead highlight the need to treat learning outcomes and test scores as distinct policy outcomes when evaluating the effectiveness of remedial education programs.

## VI. Conclusion

An evaluation of an innovative remedial education program finds that it was both successful and cost effective in improving basic learning outcomes, despite

**TABLE 5**
**HETEROGENEOUS EFFECT ON WEAKER STUDENTS**

| | Class × Exam Scores | | | | ASER End-Line Scores | |
|---|---|---|---|---|---|---|
| | English (1) | Tamil (2) | Math (3) | Total (4) | Read (5) | Math (6) |
| Weak | −10.63*** | −13.37*** | −5.45*** | −9.65*** | −.48*** | −1.10*** |
| | (1.45) | (1.77) | (1.35) | (1.39) | (.09) | (.11) |
| Treatment | −.35 | −2.19 | 1.53 | −1.14 | .92*** | .61*** |
| | (2.64) | (2.36) | (1.98) | (1.69) | (.19) | (.17) |
| Treat × Weak | 3.97* | 5.87** | 1.19 | 3.44** | .03 | .55*** |
| | (2.18) | (2.38) | (1.55) | (1.75) | (.13) | (.16) |
| *p*-val (OLS) | .07 | .01 | .44 | .05 | .80 | .00 |
| *p*-val (RI) | .00 | .00 | .28 | .00 | .68 | .00 |
| Control mean | 51.22 | 64.56 | 45.68 | 56.77 | | |
| $R^2$ | .18 | .20 | .13 | .20 | .28 | .27 |
| Observations | 814 | 814 | 814 | 814 | 887 | 887 |

**Note.** Outcome variables in cols. 1–4 are the percentage marks scored by a student in his or her AISSE exams in English, Tamil, math, and total, respectively. Outcome variables in cols. 5 and 6 are the ASER scores for reading and math that have been standardized to have mean 0 and standard deviation 1 for the control group. Weak is a dummy variable that takes the value 1 if a student could not do simple division/multiplication at baseline and 0 otherwise. All regressions include individual and school controls. Individual controls include gender, age, religion, and caste. School controls include language of instruction and number of boys and girls. Wild-bootstrapped standard errors are clustered at the school level and reported in parentheses. *p*-val (OLS) reports the *p*-value for the Treat × Weak coefficient estimated by the wild-bootstrapped clustered standard errors, while *p*-val (RI) reports the *p*-value using the randomized inference method.

* $p < .10$.
** $p < .05$.
*** $p < .01$.

students being older and having significant foundational learning deficiencies at baseline. Understanding which components of the training program, such as technology aids, facilitators, group exercises, and so forth, were more effective is an important question that we leave for future work. In addition, the unique opportunity provided to benchmark test scores from national exams highlights important gaps in terms of these national administrative exams being able to accurately measure grade-level proficiency as opposed to rote memorization or poor implementation. In particular, it calls for rethinking the utility of these exams and how they can be designed and implemented better, especially in contexts where they are crucial in determining long-run outcomes of individuals.

## References

Andrabi, T., J. Das, and A. I. Khwaja. 2015. "Delivering Education: A Pragmatic Framework for Improving Education in Low-Income Countries." Policy Research Working Paper no. 7277, World Bank, Washington, DC.

Angrist, N., P. Bergman, C. Brewster, and M. Matsheng. 2020. "Stemming Learning Loss during the Pandemic: A Rapid Randomized Trial of a Low-Tech Intervention

in Botswana." CSAE Working Paper no. 2020-13, Centre for the Study of African Economies, University of Oxford.

Banerjee, A. V., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukherji, M. Shotland, and M. Walton. 2016. "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of 'Teaching at the Right Level' in India." NBER Working Paper no. 22746 (October), National Bureau of Economic Research, Cambridge, MA.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *Quarterly Journal of Economics* 122, no. 3:1235–64.

Battaglia, M., and L. Lebedinski. 2022. "With a Little Help from My Friends: Medium-Term Effects of a Remedial Education Program Targeting Roma Minority." *Economics of Education Review* 86:102196.

Beg, S., W. Halim, A. M. Lucas, and U. Saif. 2022. "Engaging Teachers with Technology Increased Achievement, Bypassing Teachers Did Not." *American Economic Journal: Economic Policy* 14, no. 2:61–90.

Blimpo, M. P. 2014. "Team Incentives for Education in Developing Countries: A Randomized Field Experiment in Benin." *American Economic Journal: Applied Economics* 6:90–109.

Cameron, C. A., J. B. Gelbach, and D. L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." *Review of Economics and Statistics* 90, no. 3:414–27.

Cascio, E., B. Sacerdote, D. Staiger, and M. Tine. 2024. "Report from Working Group on the Role of Standardized Test Scores in Undergraduate Admissions." Dartmouth University. https://home.dartmouth.edu/sites/home/files/2024-02/sat-undergrad-admissions.pdf.

Cilliers, J., I. M. Mbiti, and A. Zeitlin. 2021. "Can Public Rankings Improve School Performance? Evidence from a Nationwide Reform in Tanzania." *Journal of Human Resources* 56, no. 3:655–85.

Das, J., A. Singh, and A. Y. Chang. 2022. "Test Scores and Educational Opportunities: Panel Evidence from Five Developing Countries." *Journal of Public Economics* 206:104570.

de Barros, A., J. Fajardo-Gonzalez, P. Glewwe, and A. Sankar. 2020. "Learning by Doing? Experimental Evidence on Activity-Based Instruction in India." RISE Programme, University of Oxford.

Duckworth, A., and P. Quinn. 2009. "Development and Validation of the Short Grit Scale (Grits)." *Journal of Personality Assessment* 91:166–74.

Evans, D. K., and A. Popova. 2016. "What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews." *World Bank Research Observer* 31, no. 2:242–270.

Fagen, A. P., C. H. Crouch, and E. Mazur. 2009. "Peer Instruction: Results from a Range of Classrooms." *Review of Economics and Statistics* 91, no. 3:437–56.

Fisher, R. A. 1935. *The Design of Experiments*. Edinburgh: Oliver & Boyd.

Glewwe, P., and K. Muralidharan. 2016. "Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications." *Handbook*

*of the Economics of Education*, vol. 5, ed. E. A. Hanushek, S. Machin, and L. Woessmann, 653–743. Amsterdam: North-Holland.

Heß, S. 2017. "Randomization Inference with Stata: A Guide and Software." *Stata Journal* 17, no. 3:630–51.

Kingdon, G. G. 2007. "The Progress of School Education in India." *Oxford Review of Economic Policy* 23, no. 2:168–95.

Kremer, M., C. Brannen, and R. Glennerster. 2013. "The Challenge of Education and Learning in the Developing World." *Science* 340, no. 6130:297–300.

Kremer, M., E. Miguel, and R. Thornton. 2009. "Incentives to Learn." *Physics Teacher* 40:206–9.

Kumar, G. N. 2023. "Improving Public School Productivity: Evidence from Model Schools in India." *Economics of Education Review* 97:102465.

Lai, F., R. Luo, L. Zhang, X. Huang, and S. Rozelle. 2015. "Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing." *Economics of Education Review* 47:34–48.

Lee, D. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies* 3, no. 76:1071–102.

Linden, L., A. Banerjee, and E. Duflo. 2003. "Computer-Assisted Learning: Evidence from a Randomized Experiment." Poverty Action Lab Paper no. 5, Abdul Latif Jameel Poverty Action Lab, Cambridge, MA.

Marinelli, H. A., S. Berlinski, and M. Busso. 2024. "Remedial Education: Evidence from a Sequence of Experiments in Colombia." *Journal of Human Resources* 59, no. 1:141–74.

Mbiti, I., K. Muralidharan, M. Romero, Y. Schipper, C. Manda, and R. Rajani. 2019. "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania." *Quarterly Journal of Economics* 134, no. 3:1627–73.

Mbiti, I., and D. Rodriguez-Segura. 2022. "Back to Basics: Curriculum Reform and Student Learning in Tanzania." RISE Working Paper no. 22/099, Research on Improving Systems of Education, Oxford.

Muralidharan, K. 2013. "Priorities for Primary Education Policy in India's 12th Five-Year Plan." *India Policy Forum* 9, no. 1:1–61.

Muralidharan, K., and N. Prakash. 2017. "Cycling to School: Increasing Secondary School Enrollment for Girls in India." *American Economic Journal: Applied Economics* 9, no. 3:321–50.

Muralidharan, K., A. Singh, and A. J. Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review* 109, no. 4:1426–60.

Muralidharan, K., and V. Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119, no. 1:39–77.

Pratham. 2023. "Annual Status of Education Report 2022." New Delhi: Pratham.

Pritchett, L. 2013. *The Rebirth of Education: Schooling Ain't Learning*. Washington, DC: Center for Global Development.

Pritchett, L., and A. Beatty. 2012. "The Negative Consequences of Overambitious Curricula in Developing Countries." Working Paper no. 293, Center for Global Development, Washington, DC.

Saavedra, J., E. Näslund-Hadley, and M. Alfonso. 2017. "Targeted Remedial Education: Experimental Evidence from Peru." NBER Working Paper no. 23050 (January), National Bureau of Economic Research, Cambridge, MA.

Schell, J., B. Lukoff, and E. Mazur. 2013. "Catalyzing Learner Engagement Using Cutting-Edge Response Systems in Higher Education." In *Increasing Student Engagement and Retention Using Classroom Technologies: Classroom Response Systems and Mediated Discourse Technologies*, ed. C. Wankel and P. Blessinger. Leeds: Emerald. https://doi.org/10.1108/S2044-9968(2013)000006E011.

Singh, A. 2024. "Improving Administrative Data at Scale: Experimental Evidence on Digital Testing in Indian Schools." *Economic Journal* 134, no. 661:2207–23.

West, M., M. Kraft, A. Finn, and A. Duckworth. 2014. "Promise and Paradox: Measuring Students Non-cognitive Skills and the Impact of Schooling." *Educational Evaluation and Analysis* 38, no. 1:148–70.

Young, A. 2019. "Channeling Fisher: Randomization Tests and the Statistical Insignificance of Seemingly Significant Experimental Results." *Quarterly Journal of Economics* 134, no. 2:557–98.

Zhang, P., L. Ding, and E. Mazur. 2017. "Peer Instruction in Introductory Physics: A Method to Bring about Positive Changes in Students Attitudes and Beliefs." *Physical Review Physics Education Research* 13:010104.