Too little, Too late: Improving Post-primary Learning Outcomes in India*

Gaurav Chiplunkar
University of Virginia

Diva Dhar
University of Oxford

Radhika Nagesh

June 29, 2020

Abstract

We evaluate an innovative blended learning after-school intervention, which aims to improve learning outcomes and non-cognitive skills for Grade 9 students in government schools in India. The Avanti program run by trained facilitators, consists of online computer-based video lectures and facilitated, classroom-based interactive and peer learning activities. Using a randomized controlled trial, we find that over one academic year, the program translates into large improvements in basic Math and Reading scores for the beneficiary Grade 9 students, who were found to have baseline competencies at Grade 2 level, on average. However, we find no treatment effect on exam scores of a standardized national Grade 10 exam or on non-cognitive skills or parental support. We conclude that while the program was successful in improving foundational skills, it was not sufficient to overcome the extent of academic deficiencies at the post-primary level, where more advanced skills are needed to be at grade level. This underlines the need to re-think the strategy for post-primary initiatives to overcome deficiencies for academically weaker students and underlines the importance of improving learning outcomes at an earlier stage.

*We thank Avanti Fellows, especially Akshay Saxena and Deepak Kamble for partnering with us on this project. We are grateful for funding from Macarthur Foundation, especially to Dipa Nag Chowdhuri. We thank Clare Leaver, Anandi Mani and Karthik Muralidaran for their comments. Contact information: Chiplunkar: ChiplunkarG@darden.virginia.edu; Dhar: diva.dhar@bsg.ox.ac.uk (corresponding author); Nagesh: radhika.n89@gmail.com

1 Introduction

Educational investment to improve human capital development as a driver for "inclusive growth" has been the focus of education policy in India, as well as many other developing countries. As a result, India has experienced significant improvements in areas of school access, enrollment, infrastructure, pupil-teacher ratios and teacher salaries (Muralidharan (2013)). Despite this surge in enrolment, the quality of teaching and progress in learning levels remain low, thus failing to translate increased enrolment into better learning outcomes (and potentially better labor market outcomes in the future). For example, as of 2013, India achieved near-universal school enrollment rates in primary education and 85 percent enrollment in secondary education. However, among grade 8 students, 55 per cent could not read a simple English sentence and 57 per cent could not do a simple mathematical operation (Pratham (2016)). This suggests that students who fall behind may learn very little from their classroom instruction, which is usually above their learning levels (Banerjee et al. (2011); Ganimian et al. (2019)).

There is an extensive literature which provides multiple explanations for the continuing deficiencies in learning levels despite high enrollment (see Kremer et al. (2013) and Glewwe and Muralidharan (2016) for extensive reviews). Kremer et al. (2013) conclude that additional schooling inputs (such as extra-teachers, more textbooks or toilets) may not be as effective in changing learning outcomes as compared to improvements in pedagogical and remedial instruction and accountability reforms (such as performance-based incentives or short-term hires). To that effect, a growing body of research has evaluated the impact of pedagogical interventions to improve learning outcomes. These include Teaching at the Right Level (Banerjee et al. (2007, 2016)), adaptive technology-aided instruction (Ganimian et al. (2019)) and performance-based incentives for teachers (Muralidharan and Sundararaman (2011)).

While most of the literature has focused on improving learning outcomes in primary school, there is little evidence on how interventions impact learning outcomes at the post-primary or secondary level. This can be particularly important from a policy perspective in targeting education interventions across different grades. This paper reports on the experimental evaluation of one such technology-led, after-school instruction program at the secondary school level (Grade 9) in India. Avanti's after-school program was developed to incorporate technology into a remedial instruction program. Avanti trains facilitators to make effective use of technology aids (video recordings, presentations, worksheets etc.) along with peer learning and counseling to enhance students' cognitive and non-cognitive development. The incorporation of technology aids in an otherwise remedial program aims to improve the quality of education for marginalized students at minimal costs to the schools.

In this paper, we conduct an experiment to evaluate the impact of Avanti's after-school blended learning program run by trained facilitators at the secondary level (Grade 9) across 24 public schools in Chennai, India. We examine whether the program improves basic learning outcomes (as measured by ASER test scores), as well as non-cognitive outcomes (such as teamwork, decision making, communication etc), gender attitudes of students and parental investment in education. We find that the program leads to large gains in basic reading and math ASER scores (18.8 per cent and 39.5 per cent respectively). These effects are comparable to the Teaching at the Right Level evaluation (Banerjee, Banerji, Berry, Duflo, Kannan, Mukherji, Shotland and Walton (2016)) and the MindSpark program (Ganimian et al. (2019)). However, we find that these large gains in foundational learning outcomes do not translate into improvements in exams scores in a mandatory standardized national exam in Grade 10. Lastly, we also do not find any effect of the program on improving non-cognitive abilities and parental support.

Put together, we conclude that while carefully designed remedial education interventions can bring large gains in basic learning levels even at the post-primary level (Grade 9, in our case), it is too late to translate these gains into any improvements in more advanced grade-level exams which play a crucial role in determining future outcomes for students. Despite the improvements, 60 per cent of treatment students are not able to do division at Grade 10, which requires advanced skills such as calculus. This highlights the importance of policy debates in developing countries to address the mismatch between student learning and grade-level requirements (Ganimian et al. (2019)). Futhermore, it underscores the importance of the timing of implementing these programs at the primary school level. At the post-primary school level, even successful, intensive programs such as Avanti's might come "too late" to translate large gains in foundational skills into meaningful progress on school completion and advanced human capital accumulation.

The paper is organized as follows – Section 2 describes the empirical context and the experimental design. Section 3 describes the data collection and provides details on the final sample.

Section 4 discusses the results and Section 5 concludes.

2 Empirical Context

2.1 Avanti's program

Avanti's approach combines trained facilitators, technology aids and parent counseling, along with an emphasis on peer learning and team work.[1] 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 Maths and Science in government secondary schools in India, using basic technology and infrastructure available or installed in schools, such as computers (without internet).

Conducted after-school for an hour for five days a week, Avanti classes run for 40 weeks in a single 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 STEM, and do not having teaching certification or qualifications.They are trained on facilitating classroom sessions which rely on peer learning, team work and student collaboration. The facilitators rely on video lectures, presentations, worksheets and other academic materials developed by Avanti for Math and Science in the regional language and context. In the study schools, the program materials and discussions were in Tamil, the local language in Chennai, although the schools themselves are both English and Tamil medium schools. Finally, the program also helped provide occasional career counseling and guidance for students and parents.

The main elements of Avanti's adapted collaborative blended learning model for India thus are - a) technology aids (offline videos, worksheets etc.) in the local language b) facilitator-led classroom instruction (c) peer learning through team exercises and discussion. Table 1 provides a summary of the principal components of the Avanti program.

[1] Peer learning for example, engages students during class through worksheets and multiple choice questions (ConcepTests), where they are organized into learning groups based on their academic achievement levels and collaboratively encouraged to solve the worksheets. Facilitators move from group to group, encouraging discussion and resolving queries when necessary.

2.2 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. Of the total 281 schools under the CMC, data on grade 9 students for the academic years 2014-15 and 2015-16 was provided by the Education Department of the Corporation of Chennai. Schools were selected based on the following eligibility criteria: a) a minimum of 45 students in grade 9[2]; b) a maximum of 110 students[3]; c) 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 Chennai Municipal Corporation, 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 in the treatment schools. From each sample school, 45 students from each school were randomly sampled based on student registers at each school for a survey at the beginning of the academic year (baseline). However, out of 1080 students selected from student registers for the baseline, 46 students (approximately 4 per cent) were not available. They were not surveyed either due to 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 enrolment and the baseline. Our study sample therefore comprises of 1,044 students at baseline, who were followed up with at endline. However, during the year, some students dropped out or transferred to other schools resulting in a final (endline) sample of 880 students (an attrition of 15.7 per cent).

3 Data

We implemented three sets of student surveys at baseline and endline. The first survey included questions on students' demographic and household information, academic performance, aspirations and expenditure on education (books, tuition etc.). Second, we conducted a battery of tests to measure gender attitudes and non-cognitive skills such as critical thinking, communication, goal-

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

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

setting, problem solving, grit, self-esteem and team work (West et al. (2014); Duckworth and Quinn (2009)). The survey tools were administered in the local language (Tamil). Appendix B describes the survey questions as well as provides the construction of the composite scores used in the analysis. Third, we measured learning outcomes through the standardized ASER testing tool in Reading and Mathematics. The ASER test uses a five-point grading system for both reading and math through a survey that is administered to school students twice over the course of one year. The ASER instrument has been widely adopted as a universal tool to assess students' learning levels at different stages of education, testing their mastery of foundational skills, aligned with up to Grade 2 level reading skills and up to Grade 4 level mathematics ability. The same test is administered to all students between the ages of 5 and 16, thus the tool offers an easy baseline with which to judge basic literacy of students across grades. The test is administered orally, individually, and last about ten minutes[4]. The standard ASER implementation methodology was applied in our study. In addition to the ASER scores, we also collected exam scores for all subjects in the national standardized grade 10 examination in 2018 and match them to students in the sample. Lastly, we conducted a short survey with students' parents to better understand parental involvement in, attitude towards, and support for the child's education, measured as cumulative category scores using a series of composite Likert scales.

3.1 Sample description

Panel A of Table 3 describes the characteristics of the 1,044 students in 24 schools in our baseline sample in column (2). 40.6 per cent of students are female, and students are on average 14-15 years of age (which is appropriate for grade 9 students). 77.5 per cent of students come from socially disadvantaged backgrounds (scheduled castes, scheduled tribes and other backward castes). 26.5 per cent of students have parents who are illiterate and 56.2 per cent of households earn a monthly income less than Rs. 10,000 (approximately 550 USD adjusted for purchasing power). Columns (3) and (4) of table 3 then reports the average separately for students in the control and treatment schools, while column (5) reports the difference and its statistical significance at conventional levels. The balance test shows that the treatment group has a slightly higher proportion of students who

[4] Information on the modalities of conducting the ASER test and tools is retrieved from aser-centre.org/Survey/Basic/Pack/Sampling/History/p/54.html

are Hindu or whose guardians' income is below Rs 10,000. Below, we discuss how we control for imbalances at baseline.

Panel B describes the cognitive, non-cognitive ability of students as well as the parental support they receive. As described before, the ASER tool is used to measure basic cognitive ability. The scores indicate that on averag, students are able to read words and short sentences, but cannot read out paragraphs. On average, they have mastered addition/subtraction, but not multiplication/division. This is indicative of the massive learning gap for Grade 9 students in these schools at baseline itself. The scores indicate that the average Grade 9 students have only mastered skills usually taught in Grades 2 and 3, and are lagging far behind the grade level skills needed. This finding resonated with other studies in the Indian context which document the wide learning gap. On the balance tests, from column (5), it is important to note that treatment students perform better on cognitive tests for Reading and Math than control students at baseline - on average, they are ahead by one level on the ASER test. However, the students in control receive more parental support, supervision and investment in their education.

Given the small size of the sample, it is not surprising to find imbalance in a few baseline characteristics. However, these characteristics could be important in determining final outcomes, and we control for these baseline covariates to ensure the robustness of our results. We address the imbalance by following Bruhn and McKenzie (2009) and controlling for baseline outcome variables in all our specifications as elaborated on in the subsequent section. Lastly, the measures of life skills and gender attitudes are aggregated into an index (see appendix B for details) and appear to be balanced across treatment and control.

4 Results

We now turn to examining the effect of the intervention for an individual i in school s. We estimate the following specification:

$$Y_{isE} = \alpha + \beta T_s + \gamma_1 Y_{isB} + \delta_1 X_i + \delta_2 X_s + \varepsilon_{is}$$

where Y_{isE} and Y_{isB} are the endline and baseline outcome variables of the individual, T_s is the treatment indicator, X_i and X_s are time-invariant characteristics of the individual and school respectively. We use age, gender, religion, dummy for caste (SC/ST, OBC and others), educational qualification of the guardian and household income for X_i and number of boys and girls, indicator of a co-education school and language of instruction for X_s. We estimate the above specification both without and with individual and school controls and report them in panels A and B in each table respectively. Lastly, since the randomization was done with 24 schools, we wild-bootstrap cluster our standard errors at the school level, as suggested by Cameron et al. (2008) for statistical inference with small clusters. This is reported in all tables as the p-val (OLS) value. Furthermore, we also report the p-value from a two-sided randomization inference test (p-val (RI)). This test, originally proposed by Fisher (1935) and developed by Young (2019); Heß (2017) allows statistical inference by comparing the realized treatment effect with multiple (500) placebo assignments. This procedure therefore has the advantage of providing inference with correct size, regardless of the sample and cluster size.

4.1 Impact on cognitive ability

We begin by examining the effect of the treatment on cognitive ability as measured by the ASER reading and math scores. Both the reading and math scores can take a value from 1 to 5 as described earlier. Columns (1) and (2) of Table 4 report the results. Panel A reports the results without any individual and school controls, whereas Panel B reports the results after controlling for individual and school characteristics. We make two observations before discussing the results. First, baseline and endline variables are strongly positively correlated. Second, controlling for individual and school characteristics in addition, does not affect the magnitude and significance of the treatment coefficient a lot. Turning to the impact of the treatment in Panel B (our preferred specification), after controlling for baseline scores, along with individual and school charactersitics, students in treatment schools improve their ASER reading and math scores by 0.56 and 1.1 levels respectively, in comparison to the control group at the endline. This is equivalent to a 19 and 39.5 per cent gain in the reading and math scores respectively vis-a-vis the control group. In standard deviations, the effect size is 0.52 SD in reading and 0.86 SD in Math. This impact compares more than favorably with similar programs evaluated in India using ASER such as Pratham's Read India campaign (Banerjee et al. (2016)) and Mindspark program (Ganimian et al. (2019)).

Lastly, since the ASER scores are ordinal variables with a finite range, we re-estimate the above specification using an ordinal logit specification and find that the results are qualitatively similar for both outcome variables.

4.2 Impact on non-cognitive skills and parental support

Apart from examining the impact on cognitive ability, we also examine how the intervention affected non-cognitive skills, gender attitudes and parental support to the child. We conduct a battery of standard tests to capture non-cognitive skills and gender attitudes along with survey data to understand the parental role in helping their child in school (see appendix B for details). Given the richness of this data collection process, the primary empirical challenge is that a large set of outcomes could be potentially affected by the intervention. The richness of the data implies a danger of 'cherry-picking' outcomes that show large treatment effects. We deal with these systematically in three ways: first, we aggregate all components of a family of outcomes to form an aggregate index, such as for grit or for decision-making. Second, we undertake a principal component analysis to use the underlying variation that drives these family of outcomes. Third, we follow Banerjee et al. (2010); Kling et al. (2007) and calculate the average standardized effect over the family of outcomes. For example, for a family with N different outcomes, each denoted by n, the average effect of the treatment $\hat{\beta}$ is calculated as:

$$\hat{\beta} = \frac{1}{N} \sum_{n=1}^{N} \frac{\hat{\beta}_n}{\hat{\sigma}_n}$$

where $\hat{\sigma}_n$ is the standard deviation of the control group for outcome n. The system across all N outcomes is estimated using seemingly unrelated regression models to account for correlation among the coefficients for all outcomes in one family and the variance-covariance system is used to calculate the standard error of the estimate. We consider three family of outcomes described below.

First, we examine the impact of our intervention on parental support. This consists of components (evaluated on a scale of 1-5) on how involved parents were with school events, homework, participating in school activities and talking to the child about school. We also examine the expenditure of the household on school material (such as books etc.). Columns (1) and (2) of table

5 show the effect on the index (aggregated across all components) while column (5) reports from using the principal component analysis. Like before, panels A and B provide results without and with the individual and school controls respectively. Column (1) in table 6 reports the average standardized effect. Lastly, table A1 in the appendix A reports on the impact of the intervention on each component of the parent support index. We see robustly across all methods that there is no effect of the treatment on parental support to the child. Even the estimated coefficients are very small in magnitude (as compared to say, the mean in the control group). The statistical inference is also robust to using the randomization inference test, as reported by the p-val (RI) values in the table.

Second, we conduct a set of standard non-cognitive skill tests. These 'life-skills' include communication skills, critical thinking, decision making, goal setting, grit and problem solving. Each metric is measured on a scale of 1 to 5 on a number of questions. Similar to the previous case, we report the index (aggregated across all components) in column (3) of table 5, the principal component in column (6) of the same table and the average standardized effect in column (2) of table 6. Furthermore, we also report the impact of the intervention on each component separately (instead of the index) in table A2 in appendix A. Again, we find no robust impact of the intervention on any measure of non-cognitive skills.

Lastly, we turn to examining whether the intervention had an impact on student attitudes towards gender. We ask students on how strongly they agree or disagree with respect to gender roles, equality in opportunities, higher studies, opportunities to boys and education of a girl (see appendix B for detailed questions). We report the index (aggregated across all components) in column (4) or table 5, the principal component in column (7) of the same table and the average standardized effect in column (3) of table 6. Lastly, we also report the impact of the intervention on each component separately in table A3 in appendix A. We do not find a systematic impact of the intervention on gender attitudes.

To briefly summarize, we find that the intervention does have large and robust impacts in improving the cognitive ability of students, but does nothing to improve their non-cognitive abilities, gender attitudes or parental support.

4.3 Impact on standardized national exam scores

In class 10, all students in India write a standardized national exam. Given the large improvement in cognitive ability, we turn to examining whether this translates to improvement in the scores that students get on this national exam. We make a minor modification to the regression specification to control for baseline cognitive ability differences across students, apart from the individual and school controls. We examine the impact across all five subjects (Maths, English, Tamil, Science and Social Science) as well as the total exam score. Each score is normalized to have mean 0 and standard deviation 1 in the control group. The results are reported without individual and school controls in panel A of table 7 and with these controls in panel B. We see that baseline ASER reading and math scores are strongly positively correlated with the exam scores, but after controlling for them, the treatment does not have any significant positive impact on test scores. In fact, the estimated coefficients are negative and comparable in magnitude to the cognitive ability scores. Moreover, the statistical inference is robust to using randomizaiton inference tests, as reported by the p-val (RI) values. The above finding suggests what we want to highlight in our paper – the intervention does robustly improve basic cognitive ability, but fails to improve non-cognitive ability and fails to translate into significant gains in a standardized national exam one year later.

4.4 Attrition

As mentioned previously, we had 1044 students in the baseline sample at the beginning of the academic year, but were only able to measure outcomes for 880 students at the end of the academic year (an attrition of 15.7 per cent). This attrition is caused both due to students dropping out of school as well as transfers to other schools, for example, due to family migration. In section C in the appendix, we check for differential attrition between control and treatment groups, and see if this affects our estimates of treatment effects. Specifically, we do so in two ways: first, we examine whether student and parent characteristics, cognitive and non-cognitive ability and parental support are differentially correlated with dropouts across treatment and control schools. 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). These bounded estimates are consistent with our main analysis. Table C2 shows that the attrition-adjusted lower bound on the point estimate is 0.27 for Reading and 0.63 for Math. We therefore conclude that the attrition is not differential across control and treatment schools and does not change the conclusions we draw from the study.

5 Conclusion

This paper examines whether a post-primary program led by facilitators (rather than teachers) with technological inputs, can succeed in improving learning outcomes, team work and non-cognitive skills for students. We find that the intervention achieves improved learning outcomes in basic cognitive ability. However, it does not translate into improvements in exam scores across disciplines in national standardized Grade 10 exams. While the results of the program are encouraging in terms of shifting much needed basic literacy competencies, this program may be an example of "too little, too late", indicating the urgency to address learning levels more urgently at the primary level itself.

Directions and questions for further research include understanding and unpacking which elements of the program were most effective. While this was not possible in our research design given the combination of elements in the Avanti program model, there is value in disentangling the contribution of individual elements of the program. Moreover, given the limited duration of the pilot program, future research should aim to capture the impact on longer exposure to programs which aim to combine peer learning and technology. It would also be valuable to understand which grade level Avanti should address their program, given that many students are in need of improved reading and Math scores, which need to be improved before they can achieve gains in more advanced tests such as standardized exams.

References

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukherji, Marc Shotland, and Michael Walton, "Mainstreaming an effective intervention: Evidence from randomized evaluations of Teaching at the Right Level in India," Technical Report, National Bureau of Economic Research 2016.

Banerjee, Abhijit V, Abhijit Banerjee, and Esther Duflo, *Poor economics: A radical rethinking of the way to fight global poverty*, Public Affairs, 2011.

_ , Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani, "Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India," *American Economic Journal: Economic Policy*, 2010, *2* (1), 1–30.

_ , Shawn Cole, Esther Duflo, and Leigh Linden, "Remedying education: Evidence from two randomized experiments in India," *The Quarterly Journal of Economics*, 2007, *122* (3), 1235–1264.

Bruhn, Miriam and David McKenzie, "In pursuit of balance: Randomization in practice in development field experiments," *American economic journal: applied economics*, 2009, *1* (4), 200–232.

Cameron, A Colin, Jonah B Gelbach, and Douglas L Miller, "Bootstrap-based improvements for inference with clustered errors," *The Review of Economics and Statistics*, 2008, *90* (3), 414–427.

Duckworth, A. L. and P. D. Quinn, "Development and validation of the Short Grit Scale (GritS)," *Journal of Personality Assessment*, 2009, *91*, 166–174.

Fagen, Adam P., Catherine H. Crouch, and Eric Mazur, "Peer Instruction: Results from a Range of Classrooms," *The Review of Economics and Statistics*, 2009, *91* (3), 437–456.

Fisher, Ronald A, "The Design of Experiments," 1935.

Ganimian, Alejandro J, Karthik Muralidharan, and Abhijeet Singh, "Disrupting education? Experimental evidence on technology-aided instruction in India," *American Economic Review*, 2019, *109* (4), 1426–60.

Glewwe, Paul and Karthik Muralidharan, "Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications," *Handbook of the Economics of Education*, 2016, *5*, 653–743.

Heß, Simon, "Randomization inference with Stata: A guide and software," *The Stata Journal*, 2017, *17* (3), 630–651.

Kling, Jeffrey R, Jeffrey B Liebman, and Lawrence F Katz, "Experimental analysis of neighborhood effects," *Econometrica*, 2007, *75* (1), 83–119.

Kremer, Michael, Conner Brannen, and Rachel Glennerster, "The Challenge of Education and Learning in the Developing World," *Science*, 2013, *340* (6130), 297–300.

Lee, David, "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *Review of Economic Studies*, 2009, *3* (76), 1071–1102.

Muralidharan, Karthik, "Priorities for primary education policy in Indias 12th five-year plan," *India Policy Forum*, 2013, *9* (1), 1–61.

_ **and Venkatesh Sundararaman**, "Teacher Performance Pay: Experimental Evidence from India," *Journal of Political Economy*, 2011.

Pratham, "Annual Status of Education Report," Technical Report, Pratham 2016.

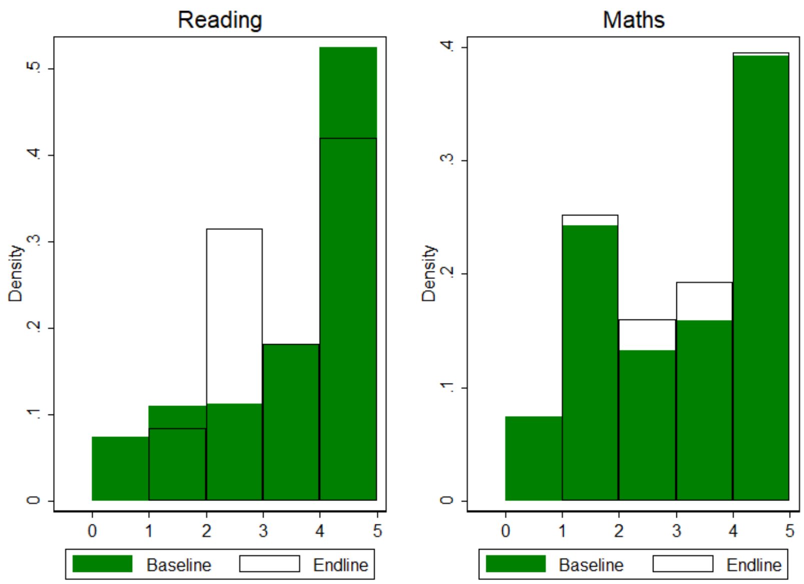
Schell, Julie, Brian Lukoff, and Eric Mazur, "Catalyzing Learner Engagement Using Cutting-Edge Response Systems in Higher Education," 2013.

West, Martin, Matthew Kraft, Amy Finn, and Angela Duckworth, "Promise and Paradox: Measuring Studentsâ Non-cognitive Skills and the Impact of Schooling," 2014.

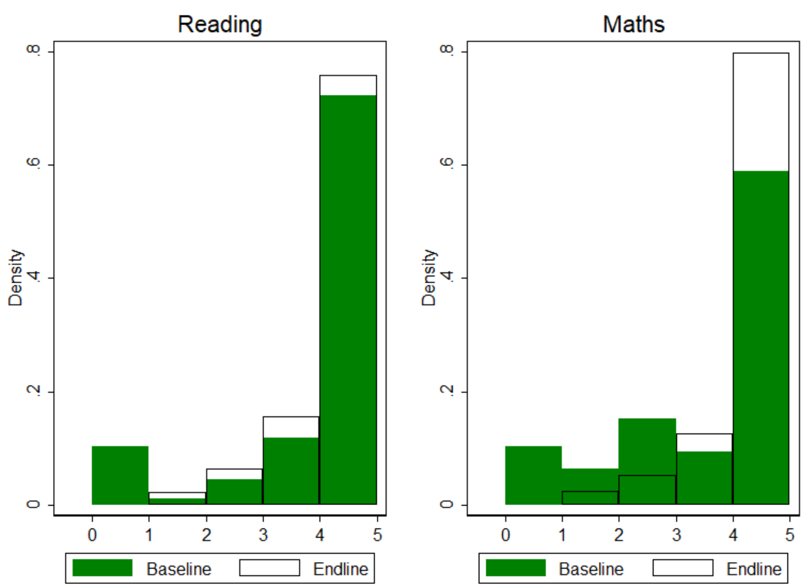
Young, Alwyn, "Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results," *The Quarterly Journal of Economics*, 2019, *134* (2), 557–598.

Zhang, Ping, Lin Ding, and Eric Mazur, "Peer Instruction in introductory physics: A method to bring about positive changes in students attitudes and beliefs," *Physical Review Physics Education Research*, 2017.

6 Figures and tables



(a) Control schools



(b) Treatment schools

Figure 1: Reading and math scores in control and treatment schools

Table 1: Avanti Program Components Description

Technology: 15 short five-minute videos in the local language that explain concepts through the use of real world examples.	1 hour 15 minutes (15 videos) per week
Peer-learning: Students organized in groups of 6 and tasked with collaboratively solving worksheets. Facilitators intervene to encourage discussion and resolve questions only when necessary.	1 hours 30 minutes (6 worksheets) per week
Trained Facilitators: Facilitators (from the local area) led lectures to summarize key concepts based on facts emerging from an in-class discussion.	1 hour 40 minutes (20 min per day) per week

Table 2: ASER Scores Description

Reading Scores	**Maths Scores**
Level 1: Reading of letter	Level 1: Identifying 1/2 digit numbers
Level 2: Reading of word	Level 2: Identifying 3 digit numbers
Level 3: Reading of sentence	Level 3: Addition / Subtraction (2 digit)
Level 4: Reading of paragraph	Level 4: Multiplication (2 digit x 1 digit)
Level 5: Reading of Story	Level 5: Division (2 digit by 1 digit)

Table 3: Sample characteristics

	Whole sample		Control (C)	Treatment (T)	Difference
	N	Mean/SD	Mean/SD	Mean/SD	T - C
	(1)	(2)	(3)	(4)	(5)
Panel A: Student characteristics					
Female	1044	0.406	0.400	0.413	0.013
		[0.015]	[0.021]	[0.022]	
Age	1043	14.820	14.789	14.851	0.061
		[0.026]	[0.037]	[0.036]	
Hindu	1044	0.733	0.706	0.760	0.053*
		[0.014]	[0.020]	[0.019]	
SC/ST	1044	0.366	0.348	0.384	0.035
		[0.015]	[0.021]	[0.021]	
OBC	1044	0.409	0.405	0.413	0.007
		[0.015]	[0.021]	[0.022]	
Guardian characteristics:					
Illiterate	1044	0.265	0.263	0.267	0.004
		[0.014]	[0.019]	[0.020]	
Any schooling	1044	0.711	0.716	0.705	-0.010
		[0.014]	[0.020]	[0.020]	
Income <Rs.10,000	1044	0.562	0.515	0.610	0.095***
		[0.015]	[0.022]	[0.021]	
Income >Rs.10,000	1044	0.120	0.123	0.116	-0.007
		[0.010]	[0.014]	[0.014]	
Panel B: Student cognitive, non-cognitive ability and parental support					
ASER Reading	954	3.820	3.299	4.368	-1.069***
		[0.039]	[0.052]	[0.045]	
ASER Maths	954	3.270	2.812	3.753	-0.941***
		[0.046]	[0.060]	[0.062]	
Lifeskills Index	1044	-0.019	-0.020	-0.017	-0.003
		[0.055]	[0.074]	[0.082]	
Student Attitude Index	1044	0.096	0.104	0.088	0.016
		[0.042]	[0.059]	[0.060]	
Parental Support Index	1044	14.816	15.061	14.566	0.495**
		[0.124]	[0.180]	[0.169]	
Education exptd. (rupees)	948	639.800	691.144	587.581	103.564**
		[23.702]	[35.481]	[31.216]	

Notes: See Appendix B for measurement of non-cognitive ability. Column (5) reports the difference between treatment and control and the asteriks report a t-test of whether this difference is statistically significant. * denotes significance at 0.1 level, ** at 0.05 and *** at 0.01 level.

Table 4: Impact on cognitive ability

	ASER Score		ASER Score	
	Reading	Maths	Reading	Maths
	(1)	(2)	(3)	(4)
	OLS	OLS	O. Logit	O. Logit
PANEL A: Without individual and school controls				
Treatment	0.598***	1.066***	1.331***	2.200***
	(0.141)	(0.197)	(0.325)	(0.409)
Baseline	0.614***	0.530***	1.405***	1.109***
	(0.0590)	(0.0544)	(0.165)	(0.141)
p-val (OLS)	0.00	0.00	0.00	0.00
p-val (RI)	0.00	0.00	0.00	0.00
R^2	0.506	0.537		
PANEL B: With individual and school controls				
Treatment	0.559***	1.106***	1.301***	2.430***
	(0.152)	(0.244)	(0.378)	(0.606)
Baseline	0.610***	0.508***	1.433***	1.099***
	(0.0570)	(0.0661)	(0.167)	(0.155)
p-val (OLS)	0.00	0.00	0.00	0.00
p-val (RI)	0.00	0.00	0.00	0.00
R^2	0.522	0.561		
N	695	695	695	695
Control mean	2.967	2.778	2.967	2.778

Notes: Individual controls include gender, age, religion, caste and guardian's education and income. School controls include language of instruction, size of class and an indicator variable for whether the school is co-educational. '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 randomized inference method. * denotes significance at 0.1, ** at 0.05 and *** at 0.01 respectively.

Table 5: Impact on non-cognitive ability and parental support

	Exptd.	Parent Support	Life Skills	Student Attitudes	Parent Support	Life Skills	Student Attitudes
	Total	Index	Index	Index	PCA	PCA	PCA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PANEL A: Without individual and school controls							
Treatment	25.17	0.281	0.138	-0.0250	0.0925	0.0916	0.179
	(66.90)	(0.346)	(0.166)	(0.280)	(0.115)	(0.125)	(0.123)
Baseline	0.438***	0.304***	0.225***	0.221***	0.306***	0.233***	0.343***
	(0.0841)	(0.0332)	(0.0368)	(0.0458)	(0.0311)	(0.0393)	(0.0203)
p-val (OLS)	0.707	0.417	0.405	0.929	0.420	0.465	0.146
p-val (RI)	0.726	0.488	0.422	0.940	0.510	0.482	0.196
R^2	0.083	0.084	0.050	0.032	0.086	0.053	0.110
PANEL B: With individual and school controls							
Treatment	45.43	0.0797	0.108	-0.103	0.0329	0.0673	0.187
	(88.33)	(0.344)	(0.213)	(0.337)	(0.132)	(0.157)	(0.136)
Baseline	0.442***	0.298***	0.220***	0.169***	0.299***	0.228***	0.291***
	(0.0675)	(0.0330)	(0.0327)	(0.0504)	(0.0337)	(0.0350)	(0.0259)
p-val (OLS)	0.607	0.817	0.611	0.760	0.804	0.669	0.168
p-val (RI)	0.556	0.828	0.608	0.732	0.766	0.654	0.152
R^2	0.095	0.128	0.066	0.112	0.127	0.069	0.158
N	880	880	880	899	880	880	880
Control mean	632.3	14.09	22.93	14.97	-0.130	-0.0366	-0.188

Notes: Impact on index reported in columns (1) to (4) while principal component of each composite reported in columns (5) to (8). Individual controls include gender, age, religion, caste and guardian's education and income. School controls include language of instruction, size of class and an indicator variable for whether the school is co-educational. '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 randomized inference method. * denotes significance at 0.1, ** at 0.05 and *** at 0.01 respectively.

Table 6: Non-Cognitive Skills: Average Standardized Effects

	Life Skills	Student Att	Parent Support
	(1)	(2)	(3)
Mean	0.0440	0.0484	0.0449
Standard Deviation	0.0371	0.0299	0.0386
N	880	880	880
Ind. Controls	Yes	Yes	Yes
Sch. Controls	Yes	Yes	Yes

Notes: All outcomes presented in this table are aggregate measures composed of relevant survey questions. Details are presented in Appendix 2.

Table 7: Impact on standardized national class X exam

Z-scores:	Tamil	English	Maths	Science	Social Science	Total
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Without individual and school controls						
Treatment	-0.493***	-0.278	-0.0356	-0.522**	-0.452**	-0.451**
	(0.184)	(0.211)	(0.287)	(0.205)	(0.185)	(0.198)
Base Read	0.362***	0.225***	0.0498	0.290***	0.258***	0.301***
	(0.0636)	(0.0565)	(0.0901)	(0.0481)	(0.0586)	(0.0698)
Base Math	0.191***	0.242***	0.178***	0.168***	0.170***	0.231***
	(0.0341)	(0.0264)	(0.0340)	(0.0379)	(0.0277)	(0.0319)
p-val (OLS)	0.007	0.188	0.901	0.011	0.015	0.023
p-val (RI)	0.004	0.168	0.868	0.006	0.018	0.006
R^2	0.232	0.188	0.057	0.144	0.142	0.208
PANEL B: With individual and school controls						
Treatment	-0.373**	-0.311	-0.0145	-0.205	-0.184	-0.278*
	(0.147)	(0.219)	(0.231)	(0.183)	(0.172)	(0.150)
Base Read	0.287***	0.182***	-0.0153	0.203***	0.183***	0.218***
	(0.0510)	(0.0458)	(0.0833)	(0.0453)	(0.0436)	(0.0577)
Base Math	0.186***	0.220***	0.176***	0.162***	0.175***	0.224***
	(0.0394)	(0.0271)	(0.0372)	(0.0414)	(0.0238)	(0.0321)
p-val (OLS)	0.011	0.155	0.95	0.261	0.286	0.064
p-val (RI)	0.002	0.142	0.942	0.236	0.224	0.056
R^2	0.320	0.257	0.145	0.241	0.225	0.289
N	755	748	755	755	754	756

Notes: (i) Variables 'Base Read' and 'Base Math' control for the students' baseline scores on ASER Reading and Maths. Individual controls include gender, age, religion, caste and guardian's education and income. School controls include language of instruction, size of class and an indicator variable for whether the school is co-educational. '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 randomized inference method. * denotes significance at 0.1, ** at 0.05 and *** at 0.01 respectively.

A Appendix Tables

Table A1: Impact on Parental Support

	Events	Homework	School	Talk
	(1)	(2)	(3)	(4)
PANEL A: Without individual and school controls				
Treatment	0.0820	0.0349	0.000681	0.107
	(0.133)	(0.116)	(0.164)	(0.139)
Baseline	0.256***	0.212***	0.101***	0.247***
	(0.0314)	(0.0329)	(0.0289)	(0.0315)
p-val (OLS)	0.539	0.763	0.997	0.440
p-val (RI)	0.524	0.776	0.994	0.448
R^2	0.059	0.040	0.009	0.057
PANEL B: With individual and school controls				
Treatment	0.0182	0.0710	-0.0833	0.0356
	(0.134)	(0.140)	(0.155)	(0.185)
Baseline	0.258***	0.209***	0.0947***	0.238***
	(0.0336)	(0.0308)	(0.0260)	(0.0313)
p-val (OLS)	0.892	0.613	0.590	0.847
p-val (RI)	0.884	0.572	0.626	0.804
R^2	0.088	0.053	0.044	0.082
N	880	880	880	880
Control mean	3.504	3.985	3.314	3.288

Notes: Individual controls include gender, age, religion, caste and guardian's education and income. School controls include language of instruction, size of class and an indicator variable for whether the school is co-educational. '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 randomized inference method. * denotes significance at 0.1, ** at 0.05 and *** at 0.01 respectively.

Table A2: Impact on Non-Cognitive Life-skills

	Communication	Critical Thinking	Decision	Goal Setting	Grit	Problem Solving
	(1)	(2)	(3)	(4)	(5)	(6)
PANEL A: Without individual and school controls						
Treat	0.0110	0.0357	-0.0114	-0.0336	0.0591	0.0737**
	(0.0373)	(0.0493)	(0.0485)	(0.0441)	(0.0419)	(0.0356)
Baseline	0.191***	0.240***	0.198***	0.139***	0.127***	0.148***
	(0.0323)	(0.0292)	(0.0376)	(0.0348)	(0.0380)	(0.0431)
p-val (OLS)	0.768	0.469	0.814	0.447	0.159	0.038
p-val (RI)	0.744	0.532	0.802	0.498	0.152	0.074
R^2	0.033	0.056	0.034	0.020	0.018	0.027
PANEL B: With individual and school controls						
Treat	0.0132	0.0179	-0.00608	-0.0580	0.0581	0.0826
	(0.0395)	(0.0673)	(0.0555)	(0.0531)	(0.0411)	(0.0531)
Baseline	0.187***	0.243***	0.187***	0.135***	0.122***	0.146***
	(0.0313)	(0.0260)	(0.0403)	(0.0397)	(0.0328)	(0.0406)
p-val (OLS)	0.739	0.790	0.913	0.275	0.158	0.12
p-val (RI)	0.770	0.756	0.916	0.266	0.244	0.072
R^2	0.044	0.078	0.052	0.042	0.035	0.041
N	880	880	880	880	880	880
Control mean	3.782	3.873	3.989	3.983	3.483	3.821

Notes: Individual controls include gender, age, religion, caste and guardian's education and income. School controls include language of instruction, size of class and an indicator variable for whether the school is co-educational. '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 randomized inference method. * denotes significance at 0.1, ** at 0.05 and *** at 0.01 respectively.

Table A3: Impact on Gender Attitudes

	Roles	Equal Opp.	Study	Boys Opp.	Wife Education
	(1)	(2)	(3)	(4)	(5)
PANEL A: Without individual and school controls					
Treat	0.0106	-0.0531	-0.0100	0.269***	0.0895
	(0.119)	(0.0764)	(0.0605)	(0.0927)	(0.0924)
Baseline	0.172***	0.0894***	0.210***	0.194***	0.235***
	(0.0301)	(0.0341)	(0.0459)	(0.0400)	(0.0286)
p-val (OLS)	0.929	0.487	0.868	0.00400	0.333
p-val (RI)	0.912	0.476	0.896	0.00600	0.346
R^2	0.031	0.008	0.043	0.048	0.064
PANEL B: With individual and school controls					
Treat	0.0360	-0.0792	-0.0184	0.237**	0.0946
	(0.129)	(0.0792)	(0.0771)	(0.0969)	(0.119)
Baseline	0.155***	0.0931***	0.202***	0.128***	0.199***
	(0.0305)	(0.0304)	(0.0468)	(0.0430)	(0.0298)
p-val (OLS)	0.780	0.317	0.812	0.0150	0.426
p-val (RI)	0.780	0.382	0.822	0.0180	0.326
R^2	0.076	0.032	0.056	0.112	0.107
N	880	880	880	880	880
Control mean	2.210	4.124	4.268	2.498	2.033

Notes: Individual controls include gender, age, religion, caste and guardian's education and income. School controls include language of instruction, size of class and an indicator variable for whether the school is co-educational. '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 randomized inference method. * denotes significance at 0.1, ** at 0.05 and *** at 0.01 respectively.

B Survey Questions and Indicators

Certain sets of questions seeks responses from students on a Likert Scale in the survey are analysed both in aggregate and through the principal component that retains most of the sample's information. The variable label associated with each composite question in the analysis is also included. The following questions are included in the Student Questionnaire, and measured on a Likert Scale ranging from Never (1) to Always (5). For each of the following sets of questions, the impact of treatment is observed on individual metrics, cumulative scores, as well as the principal component.

Table B1: Parent Support and Student Attitude Questions

Variable	**Survey Question**
Panel A: Parental Support	
School	How often do your parents help you with your school work?
Talk	How often do your parents talk to you about what you are doing in school?
Homework	How often do your parents ask you about homework?
Events	How often do your parents go to meetings or events at school?
Panel B: Student attitudes towards gender	
Roles	A woman's most important role is to take care of her home, feed kids and cook for her family.
Equal Opp.	Men and women should get equal opportunities in all spheres of life - education, healthcare, food, decision making.
Study	Girls should be allowed to study as far as they want.
Boys Opp.	Boys should be allowed to get more opportunities and resources for education than girls.
Education	Wives should be less educated than their husbands.

Table B2: Controls Description

	Variable Name	**Variable Description**
Individual Controls	Student Gender	Indicator; Male/Female
	Student Age	Continuous; Years
	Religion	Indicator; Hindu/Non-Hindu
	Caste	Indicators for SCST, General, and OBC castes
	Guardian Education	Categorical; Illiterate, Schooling
	Guardian Income	Categorical; <10,000, =>10,000
School Controls	Language of Instruction	Indicator; Tamil/English
	Size of class	Continuous, by gender
	Coeducational School	Indicator; Coed/Single Sex

Table B3: Life Skills Composite Questions

Variable	Survey Question
Decision Making	I look for information to help me understand the problem I consider the risk of a choice before making a decision. I think about all the information I have about the different choices I think of past choices when making new decisions
Critical Thinking	I can easily express my thoughts on a problem. I usually have more than one source of information before making a decision. I compare ideas when thinking about a topic. I keep my mind open to different ideas when planning to make a decision. I am able to tell the best way of handling a problem.
Communication	I try to keep eye contact. I recognize when two people are trying to say the same thing, but in different ways. I try to see the other person's point of view I change the way I talk to someone based on my relationship with them. I organize thoughts in my head before speaking. I make sure I understand what another person is saying before I respond.
Goal Setting	I look at the steps needed to achieve the goal. I think about how and when I want to achieve a goal. After setting a goal, I break goals down into steps so I can check my progress. Both positive and negative feedback help me work toward my goal
Problem Solving	I first figure out exactly what the problem is. I try to determine what caused the problem. I do what I have done in the past to solve the problem I compare each possible solution with the others to find the best one. After selecting a solution, I think about it for a while before putting it into action
Grit	New ideas sometimes distract me from previous ones Setbacks don't discourage me. I bounce back from disappointments faster than most people. I have been obsessed with a certain idea or project for a short time but later lost interest. I am a hard worker. I often set a goal but later choose to pursue (follow) a different one. I have difficulty maintaining (keeping) my focus on projects that take more than a few months to complete. I finish whatever I begin. I am diligent (hard working and careful).

C Attrition

Of the 1044 students in our baseline sample we only have 880 students in a the endline (an attrition of 15.7 percent). We investigate the differential attrition across treatment and control schools on observable characteristics as well as its impact on our treatment estimates in two ways. First, we report whether there is differential attrition between treatment and control. We estimate the following specification:

$$M_i = \alpha + \beta X_i + \gamma T_i + \delta T_i \times X_i + \varepsilon_i$$

where M_i is a dummy variable that takes the value 1 if the student stays in the sample and 0 if the student drops out between the baseline and endline. δ is the coefficient of interest since it tells us whether characteristics of students who stay in the sample is statistically different on average between control and treatment schools. We report the estimates for δ in table C1. As reported in the table, attrition is small in magnitude and statistically insignificant between treatment and control schools for most student characteristics.

Table C1: Differential attrition between control and treatment

Student characteristic (X_i)	$\hat{\delta}$	S.E.	p-value
	(1)	(2)	(3)
Female	-0.008	0.05	0.858
Age	0.053**	0.0271	0.05
Hindu	0.079	0.052	0.127
OBC	-0.01	0.045	0.827
Guardian Educ: Illiterate	-0.12	0.163	0.457
Guardian Educ: Schooling	-0.084	0.16	0.596
Guardian Income < Rs.10,000	-0.006	0.051	0.907
Guardian Income >= Rs.10,000	0.09	0.074	0.228
Baseline Reading ability	0.009	0.024	0.698
Baseline Maths ability	-0.025	0.019	0.17
Lifeskills (PC)	-0.0013	0.013	0.919
Gender Attitudes (PC)	-0.0231	0.018	0.204
Parent Support	-0.004	0.006	0.487
Education exptd. ('000)	0.0051	0.0032	0.109

Notes: * denotes significance at 0.1 level, ** at 0.05 and *** at 0.01 level.

In addition, to account for potential endogenous attrition from the sample, we also estimate Lee bounds on the treatment effects with bootstrapped standard errors (Lee (2009)) and report the results in table C2. Where possible, baseline values are used to tighten bounds, identified as Group 1 in table C2. In other cases (Group 2), bounds estimated using this method are tightened using gender, religion, and guardian education and income categories as covariates [5], with weighted averages of bounds defined for sub samples defined by these covariates. The table C2 shows that the attrition-adjusted upper bound on the point estimate is 0.56 for reading and 0.99 for Maths.

Table C2: Lee Bounds on Treament Effects

	Lower bound		Upper bound	
	Estimate	S.E.	Estimate	S.E.
	(1)	(2)	(3)	(4)
Group 1: Baseline Ability used to tighten bounds				
Panel A: Cognitive and Non-cognitive abilities				
Reading	0.27**	0.11	0.56***	0.11
Maths	0.63***	0.12	0.99***	0.13
Parent Support	0.073	0.34	0.30	0.34
Group 2: Individual Controls used to tighten bounds				
Lifeskills (PC)	-0.028	0.17	0.15	0.18
Gender Attitudes (PC)	0.15	0.075	0.297*	0.098
Panel B: Standardized National Exam Scores				
Tamil	-1.05	2.22	1.87	1.55
English	1.03*	1.47	3.67	1.23
Maths	1.64*	1.23	3.30	0.96
Science	-1.60	1.45	0.098	1.51
Social science	-2.23	1.77	0.28	2.01
Total	-1.18	6.82	8.98	6.65

Notes: * denotes significance at 0.1 level, ** at 0.05 and *** at 0.01 level.

[5] The variables used to tighten bounds are gender, religion, education level of the guardian and household income.