Identifying and Nurturing Math Talent: Evidence from Tamil Nadu*

Phi Adajar, Esther Duflo, Glenn Ellison, Sara Fisher Ellison, and Harini Kannan†

August 2025
PRELIMINARY: NOT FOR DISTRIBUTION

Abstract

This paper presents evidence from an RCT that used various methods to identify students from disadvantaged backgrounds in Tamil Nadu who might benefit from advanced online math courses, and offered a random sample of them high-quality online math instruction. The primary finding is that the program was able to identify a number of talented students who succeeded in the challenging coursework. The RCT was also designed to investigate the effectiveness of the various student identification methods and support features. On student identification, a predictor of engagement, obtained by distributing a course excerpt to the students beforehand, complemented administrative test scores. On student support, results indicate that tutoring did not boost engagement relative to the simpler-to-implement standalone online course.

*We thank Sadhana Babuji, Veena Jayaram, Murali Mallikarjunan, Kannan Narayanaswamy, Mustufa Patel, Akhilan Rengaswamy and the rest of the JPAL-South Asia team for their remarkable work in carrying out the project. We also have to thank the partners that made this project possible: The Art of Problem Solving, for tremendous technical and logistical assistance in addition to access to course materials, and in particular, we thank Richard Rusczyk, Yating Liu, and Jo Welsh; and The Tamil Nadu School Education Department, for their collaboration in implementing this project and in sharing administrative data. Funding was provided by the Agency Fund, Founders Pledge, and the J-PAL Science for Progress initiative. This project was approved by MIT's Committee On the Use of Humans as Experimental Subjects (Protocol Nos. 2308001072, 2408001374, and 2410001443) and the IFMR Institutional Review Board (Protocol Nos. 10705, 10708, 10713, 10740, 10775, 10776, 10795, 10822, 10854, 10916, 10987, 11013, and 11034).

†e-mail: padajar@mit.edu, eduflo@mit.edu, gellison@mit.edu, sellison@mit.edu, and harini.kannan@ifmr.ac.in

1. Introduction

The identification and nurturing of talented students provides a pathway for upward economic mobility for both the students themselves and for the society as a whole, which benefits from the contributions that talented individuals make to business and scientific research. Furthermore, students at the top of the talent distribution likely have an outsized potential in this regard.[1] Disparities exist, however, in the ability of talented students to access the necessary educational resources to attain their potential, and evidence exists that many talented students do not reach high achievement levels.[2] These disparities exist both across countries and across socioeconomic groups within countries.[3]

In this broader context, we focus on a narrower setting: we evaluate an RCT that provided online math courses to 1600 promising 7th-graders from disadvantaged backgrounds in Tamil Nadu. The RCT is part of a longer-term project that aims to identify talented 7th grade students who may benefit from advanced online coursework and other support from 7th through 12th grade. We explore three main questions. First, can we identify a substantial number of talented students and successfully encourage them to engage with and succeed in challenging online math courses? Second, what methods for identifying students are relatively more and less effective? Third, what support structures improve student engagement and performance?

Many challenges must be overcome to identify and nurture talented students from non-advantaged backgrounds. The challenges are particularly daunting in a setting like India. The limited number of spots in top universities relative to India's population creates tremendous competition for admission to those universities. For example, the Indian Institute of Technology system has expanded from the original five universities to 23 over the last thirty years, but still only 1% of all the 18-year-olds in the country are invited to take its entrance examination, the JEE Advanced. The intense competition to even *apply* leads wealthy parents to invest heavily in private schools and test preparation courses. In fact, some estimates conclude that less than 5% of students who earn qualifying scores on the JEE Advanced attended a government school, even though 60% of Indian students attend government schools.[4]

There is a substantial rural population that attend India's government school system. Rural students' schools tend to be small and far-flung, due to transportation costs. And, as a result, many

[1] See, for example, (Agarwal et al. 2024).

[2] See (Ellison and Swanson 2016) and (Banerjee et al. 2025).

[3] See (Agarwal et al. 2023; Agarwal and Gaule 2020; Ahuja et al. 2024; Steegh et al. 2019).

[4] Despite lawsuits, no official figures on the background of qualifying students are released. Anecdotal evidence from news stories touting programs designed to help government-school students on the exam suggest the fraction could be well below 5%. For example, a recent article on Punjab's PACE program indicated that 44 students from Punjabi government schools "cleared" the JEE Advanced exam in 2025. Punjab has approximately 2.5% of India's population and 48,000 students nationwide clear the JEE Advanced. The number of Delhi students who cleared the exam apparently increased from 53 in 2020 to 104 in 2024.

students attend schools which lack the scale and resources necessary to offer on-site enrichment for gifted students. Online courses offer an intriguing possibility; many of the most accomplished US math students, even those at elite schools, supplement their education with the online courses we use in our program. But there are many reasons why online math enrichment may not be feasible or effective in our non-elite Indian setting: 1) most of the students in our program have no home Internet access apart from borrowing their father's cell phone; 2) many 13-year-old students in any location will have limited interest in doing difficult mathematics outside of school; 3) it was not clear if parents would be willing to allow (let alone encourage) their children to take online math classes; 4) students might struggle with English-language content; 5) students who have fallen behind elite students over many years might struggle with advanced coursework without tutoring support; and 6) online courses can be expensive—the retail price of the course we offer in our study is approximately 20% of Tamil Nadu's average annual per capita income.

Below we report results from the first year of a planned six-year intervention. In this initial phase we identified 2490 7th grade students from 869 government schools in relatively disadvantaged parts of Tamil Nadu that we thought might benefit from or engage with an online course.[5] Students were randomly assigned at the school level to the control or a treatment in which they were given an internet-enabled tablet, access to a self-paced online math enrichment course—Prealgebra 1, offered by the Art of Problem Solving (AoPS)—and a physical course textbook. We additionally chose students for our program in eight different ways (assigned randomly across schools) to investigate the effectiveness of methods for selecting students. We also cross-randomized the treatment group into several subtreatments to investigate how a number of design features affect student engagement: tutoring assistance, the ability to work after school in a school's computer lab, and informing some students that their tablets would be taken away if they did not making sufficient progress.

The design of the full study is motivated in part by the experience of MIT's hybrid MicroMasters program in Data, Economics, and Design of Policy. The program provides high-quality instruction at low prices using online courses, and enrolls students from across the world, many from non-traditional backgrounds. The program also serves as a screening device: high performers are invited to enroll in an in-person masters degree at MIT. The students selected by this method, essentially none of whom could have gained admission to MIT otherwise, have performed very well. This non-experimental observation suggested to us that offering online courses to a larger population could be an effective way to identify talented, motivated students who are well-matched to rigorous online coursework.

Our experimental design also relates to a small, but growing literature on closing achievement gaps among gifted students. Self-paced classes and tutoring have proven useful for helping students across

[5] These students were selected from a set of 112,630 students in 3494 schools.

the ability distribution, and are potentially highly scalable (Carlana and Ferrara 2021; Koedinger et al. 1997; Muralidharan et al. 2019). Intensive programs have also been shown to affect students' long-term outcomes, such as summer programs shifting students towards elite schools (Cohodes et al. 2024). Through this experiment, we test not only the effects of long-term support and various program features, but also methods used to identify talented students for such programs in the first place.

In our context, we anticipated that many or most 13-year-olds in Tamil Nadu would not show sufficient engagement to make online coursework a good method for developing their talent. We did, however, hope that enough would be well-matched to allow us to select a cohort well suited for a multi-year program that would, by necessity, be mostly online. The selection component also featured a mini-version of the same idea: we provided a larger set of students with a printed version of one chapter of the course's textbook ("booklet") and investigated whether the degree to which students worked on the booklet over a month could help identify those who would engage with the online course.

Section 2 describes the experiment in more detail and the online course offered to students.

Section 3 provides descriptive evidence on student engagement and performance. Roughly 40% of the students engaged with most of the course content, i.e. they clicked on the links that show the readings, they started the videos, they displayed the problems, and so forth. With a stringent multifaceted definition of "successful completion"—clicking on at least 80% of course components, passing at least 90% of the unit quizzes, and getting at least 60% of the challenge problems correct—the successful completion rate is about 20%. We view this as impressive and a good first-stage result given both the level of course material relative to their school curriculum and also simply the fact that it can be hard to get 13-year-olds to engage with extracurricular math. Data on the challenge problems and our endline test suggest that the difficulty level of the AoPS material was an obstacle to wider student engagement.

Section 4 presents evidence of the relative success of students selected by different methods. Our primary finding is that using student engagement with the booklet as a means of predicting engagement with the online course is valuable. Students selected purely on the basis of their standardized test scores had the lowest average course engagement (42%), while students selected using a combination of standardized scores and booklet engagement had the highest average course engagement (58%). Other single- and combined-measure methods were between these extremes, with most statistically indistinguishable from each other. In a regression analysis of the relationship between student performance and variables available at the time of selection, booklet engagement is the strongest predictor of course engagement, with a one-standard deviation increase in booklet engagement leading to a 9 percentage point increase in engagement. Standardized test scores and

the (cheating-impacted) test we administered are weaker significant predictors of student success. Grit and teacher recommendations are not significant. Results using different metrics of student learning in place of student engagement are similar.

Section 5 presents evidence on program design. Its most striking result is that neither of our tutoring treatments boosts student completion. Indeed, we find large and highly significant *negative* effects of both the group tutoring and individual tutoring subtreatments on course completion. Students in the tutoring subtreatments have engagement that is roughly 15 percentage points lower. Results on the number of unit quizzes passed are similar. This was surprising to us given the previous literature on tutoring, especially given the difficulty of our courses and the likelihood that parents would not provide much help. Two possible explanations for the result are that time spent with tutors may crowd out time spent working on the course, and that watching AoPS videos and working on Alcumus problems may be somewhat addictive.

Section 6 presents evidence on heterogeneity in program effectiveness. The gender gap in very high STEM achievement is particularly large in India. In relation to this, it is interesting that two-thirds of the students selected for our program via a gender-blind procedure are female, and that the girls show greater course engagement and learning.

Section 7 presents evidence on student learning. We know from the unit quizzes that many students who completed the course did learn material that goes beyond what is covered in their school curriculum. Scores on our endline test, which assessed whether students could solve highly-challenging problems on material that overlaps with their school curriculum, were very low. They also again had some cheating issues. The mean non-cheating control student was only able to solve about 1.25 of the 19 challenging problems. Non-cheating treatment students solved about 1.5 problems. The 20% increase is statistically significant, albeit not as large as we had hoped. We infer in part that the difficulty level of our test was higher than it should have been, but also that we are working with a population that is farther behind the level of top US students than we had anticipated.

Our paper complements several lines of research on the economics of talent. For instance, studies have suggested that better match quality of highly-skilled and educated workers to positions could increase research output and GDP (Agarwal and Gaule 2020; Ugarov 2019). Others have documented that students of high ability often have mathematical skills that go unrecognized in the traditional school system (Banerjee et al. 2025); even conditional on academic achievement, gaps remain across gender, socioeconomic status, schools, and countries (Agarwal et al. 2023; Agarwal and Gaule 2020; Ahuja et al. 2024).

2. The Environment and Experiment

2.1. Student population Indian universities train a large number of scientists and engineers, but the population of 18-year-old Indians is also extremely large, making it challenging to earn places. India has a relatively large private school sector serving about half of the population. Multiple state governments have recently made efforts to provide pathways for talented students from government schools to prepare for positions in top universities, but the fraction of students at top universities who come from government schools appears to be very, very low.

Our study takes place in government schools in 10 districts of Tamil Nadu: Chennai, Cuddalore, Dharmapuri, Kancheepuram, Nagapattinam, The Nilgiris, Tirunelveli, Tiruvanamalai, Vellore, and Vellupuram. Although government schools generally serve a relatively disadvantaged population, the schools that were selected for our project are substantially negatively selected within the government school system. Over 90% of our schools are in rural areas (roughly half of the Tamil Nadu population is urban). Over 20% of the students we eventually select for participation in our program are members of a scheduled caste/tribe (which we will refer to as "disadvantaged"), another 45% are a member of a "most backward class" and essentially all of the rest belong to some other "backward class" and would also qualify for affirmative action programs with reserves, like those that exist for some universities and jobs.

Our experiment began with 3,494 government schools in 10 districts of Tamil Nadu, collectively enrolling 112,630 students. These schools should not be regarded as a representative sample from the full set of government schools in Tamil Nadu: we intentionally selected districts with relatively disadvantaged areas, on the basis of both socioeconomic status and the presence of scheduled castes and tribes. Further, schools were selected at a time when not all schools had reported their students' 6th-grade final exam scores, but implementation time constrants prompted us to select schools even with incomplete data.

We chose to conduct our program in 1,441 of the schools enrolling 73,533 7th-grade students, reflecting in part a preference for schools that reported having at least one very high scorer or multiple high scorers on the 6th-grade final exam. We visited each school in-person and selected a total of 9,709 initial participants that met at least one of several criteria that included sufficiently high 6th-grade test scores (in both absolute terms and relative to their school), being recommended by the student's teacher as one of the top students in the class, or being recommended by the teacher as another student who might welcome a challenge. More details about this first screen are given in the appendix.

We then selected students from this population of 9,709 high-achieving students to participate in the full RCT. To do so, each student received a printed "booklet" containing Chapter 1 of the

AoPS *Prealgebra* textbook and answer sheets to work on the problems in the text. We monitored engagement with the content using both an online form where students could check their answers, and by having teachers take photos of students' answer sheets within the booklet. After approximately 1 month, we asked students to take an online "baseline" test to assess their understanding of the concepts, and also to measure student "grit" through a small survey and a riddle designed to measure persistence. To economize on travel costs, we had schoolteachers proctor these tests (rather than sending proctors) and had students take the test on school computers, relying on the fact that by looking at patterns of matching incorrect answers on the non-multiple choice test, we would easily be able to detect students who cheated. Unfortunately, we learned that most teachers in the subject schools did not prevent their students from cheating: about 70% of the student either copied a classmate's answers or had a classmate copy their answers.

Using students' 6th-grade test scores, their booklet engagement, baseline exam scores, and grit measures, we selected 2,490 students across 867 schools to participate in our RCT; more details on selection are discussed in the treatment section.

2.2. Treatments In our main treatment, we explore the effectiveness of online courses for supporting talented students. To do so, we provided treatment students with the entire self-paced Prealgebra 1 online course, a corresponding physical textbook, an internet-connected tablet, and logistical help with accessing the course. Treated students were also placed in WhatsApp groups used to send periodic messages and connect them to other students. Treatment was randomized at the school level, and each school was treated with probability two-thirds.

To investigate how best to select students for a program like ours, we designed eight different student selection methods, adjusting parameters for each so that it would select approximately 2,400 students across the set of 3,494 schools; we then randomized which method was applied to each individual school. Comparing sample means allows us to both use sample means to estimate student engagement with the online course as a function of the procedure used. Furthermore, this procedure creates an individual-level dataset in which there is a great deal of heterogeneity in each of the characteristics that could have been used in a selection. The baseline selection method used only the data readily available to the education ministry: reported scores on the end-of-6th-grade exams. Given that some schools inflate scores, we based this selection on both students' within-school 6th-grade z-scores and their absolute score. Schools were assigned to this selection method with probability 0.1. At other schools, we selected students based on indexes using reflected information we gained from our school visits (possibly in addition to administrative scores). These included a booklet engagement index ("booklet"), baseline test score ("test"), and grit index, each of which was used with probability 0.1. The remaining schools were assigned using combinations of these criteria: booklet+test+6th grade index (20%), test+6th grade index (20%), a booklet+test

index (10%), a booklet+6th grade index (10%). The skew toward procedures incorporating the test on the material reflects that we anticipated the test would provide the best measure of the extent to which students had engaged with the chapter 1 booklets, and also did not anticipate that cheating would be so prevalent.

To investigate how program design features affect student engagement, treatment students were cross-randomized at the school level across three subtreatments as described in the table below.

Tutoring	Percent	Hi-Tech Lab	Percent	Incentive	Percent
Individual Tutoring	12.5%	Access	25%	Enforced Progress Req.	33%
Group Tutoring	37.5%	Control	75%	Unenforced Progress Req.	33%
No Tutoring	50%			No Progress Requirement	33%

The first subtreatment involved the provision of tutoring support from STEM students and graduates of elite universities in Tamil Nadu. About 12.5% of students were assigned to a one-on-one tutor. These tutors met for 15 minutes weekly with 10 different students, providing encouragement and help with any topics. Another 37.5% received group tutoring, a 1-hour meeting with approximately 20 students. The remaining 50% of students did not have a tutor. All students could also get help with material (or help other students) by posting (answering) questions on AoPS's message boards. Students receiving tutoring were placed in WhatsApp groups with their tutor. Students in the non-tutoring subtreatment were also assigned to a WhatsApp group that had 80 non-tutored students from their district intended to provide some community.

A second dimension of treatment involved how students could work on the course. In recent years, the Tamil Nadu School Education Department has invested in "Hi-Tech Labs" (HTLs) throughout government schools, providing internet-enabled computer labs for students. Each school was assigned with 25% probability to a treatment arm in which we worked with the school administration for students to work on AoPS together in the HTL at the end of the school day at least once per week. Roughly 35% of the selected schools had just a single student, so the benefit would just have be having a regular work time and a larger screen. At the remaining schools (which have 88% of the subtreatment students) students may also benefit from getting to work alongside other students.

The final subtreatment involved an incentive: some students were told that if they did not make progress in the course at a sufficient rate their tablets would be taken away. One-third of students received these messages and did eventually have the tablets taken away if they were not making enough progress. One-third received the messages but never had tablets taken away. The final third did not receive these messages. Again, these groups were randomized at the school level. In practice, the messages had not yet progressed from vague to specific by the time schools closed for summer vacation, and no tablets were collected until four months into the program; we collected tablets from approximately 100 students. Students who lost tablets could in theory earn their

tablets back by making progress on the course without the tablet, but none did.

2.3. Mathematics course Our experiment makes use of a course from the Art of Problem Solving (AoPS), a private US educational firm. AoPS began in the early 1990's with two textbooks designed for US middle school and high school students preparing for math competitions. The company has subsequently developed a nine-volume textbook series, ranging from *Prealgebra* to *Calculus*. These textbooks cover middle-school and high-school material on a subject-by-subject basis, but emphasize problem-solving and include problems that are much more challenging than the problems found in standard textbook series. AoPS offers synchronous online classes, self-paced online classes, and in-person classes. The courses serve many students preparing for math competitions and are popular with parents who want to supplement the education their children receive in their (typically elite) schools. AoPS touts that almost all of the students who have gone on to represent the US at the International Math Olympiad took its courses.

The students in our experiment were enrolled in the first of AoPS's self-paced online courses: Prealgebra 1. The course covers seven chapters of the *Prealgebra* textbook: Arithmetic, Exponents, Number Theory, Fractions, Equations and Inequalities, Decimals, and Rates, Conversions and Rates. Each chapter includes several topics (e.g., Squares, Higher Exponents, Zero as an Exponent, Negative Exponents). Each topic includes readings, sample problems, and video lectures. Most topics also contain "Alcumus" quizzes and "challenge problems," which provide a way for students to reinforce their understanding of the material. We receive data on students' engagement with each aspect of the course, and in particular, their performance on the Alcumus and challenge problems, providing us with a way to assess student learning.

The Alcumus quizzes are particularly useful for the purpose of our study. Alcumus presents questions one at a time and adjusts the difficulty level as students get answers right and wrong, targeting material to students' ability. We infer that the program presents students with questions that they have a roughly 70% chance of answering correctly. Students can aspire to "pass" or "master" each quiz by working up to and correctly answering questions of the relevant difficulty level. Some quizzes can be "passed" by correctly answering as few as two questions if the adaptive process starts with very challenging questions. Passing can take many more correct answers if the student is making mistakes. The majority of our students appear to move on once Alcumus tells them they have passed, but many of the most-engaged students persist with the quiz beyond this point in order to master topics. While these problems are difficult, the adaptive, game-like system of Alcumus provides an environment that is not only educational, but fun for many students. The Prealgebra 1 course contains 36 such quizzes.

In contrast, the course's 144 "challenge problems" are not adaptive. These problems are intended to

challenge top students, and go far beyond direct applications of the material. While many students struggle with these problems, they can also make as many attempts as they like to try to solve them.

The full course contains 366 elements: 186 lessons and videos, and the aforementioned 36 Alcumus quizzes and 144 challenge problems. Our primary measure of student engagment will be the fraction of the 366 course elements that the student has interacted with. We will assess student learning by looking at the fraction of Alcumus quizzes that they passed (or mastered) and by looking at the fraction of challenge problems that they answered correctly.

3. Program Effectiveness: Student Engagement and Learning

The most important finding of our experiment is that our student identification and support procedures were sufficient to find a substantial number of students who engaged with and learned the course material.

Figure 3.1 presents a histogram on course engagement: it records the fraction of the 366 course components clicked on by each student. The distribution is bimodal, but the most important part of the figure for our purposes is the right side: about 17% of students clicked on all 366 components of the course. Another 11% clicked on 90-99% of the components, and 8% clicked on 80-89%. There is also an equally large left tail with 29% of students engaging with less than 20% of the course. However, the hope of the intervention was always that it would identify some students who engaged and might potentially be helped by a program of additional online courses. There are a total of 578 students in the 80+% bins.

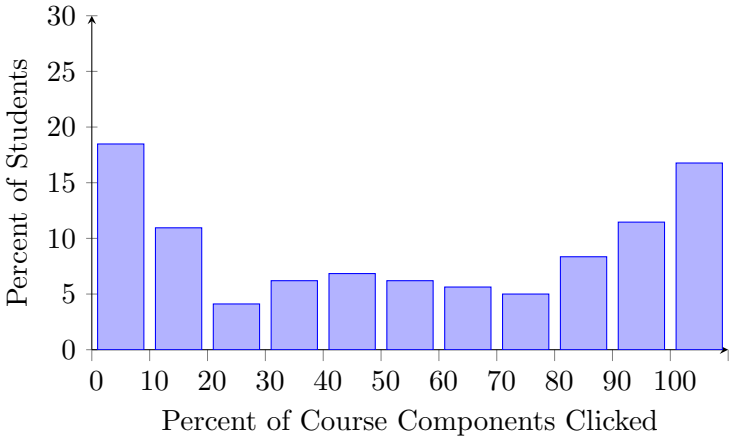


Figure 3.1: Student engagement: fraction of course components clicked

Extensive engagement with the AoPS course does not necessarily coincide with extensive learning. Some students could have clicked all the links without ever learning the material and solving any

problems.[6] Our primary source of information on whether students learned the course material is the Alcumus quizzes. Figure 3.1 presents a histogram showing the fraction of the quizzes that each student passed.[7] The fraction of students who passed all of the quizzes is 22%, which is even higher than the fraction who viewed 100% of the course components. We infer that many students not only engaged with the course but learned the material. The majority of the students who passed 100% of the Alcumus quizzes usually persisted beyond the passing mark and tried to master the material as well: the majority of those who passed all 36 quizzes mastered more than 90% of them.

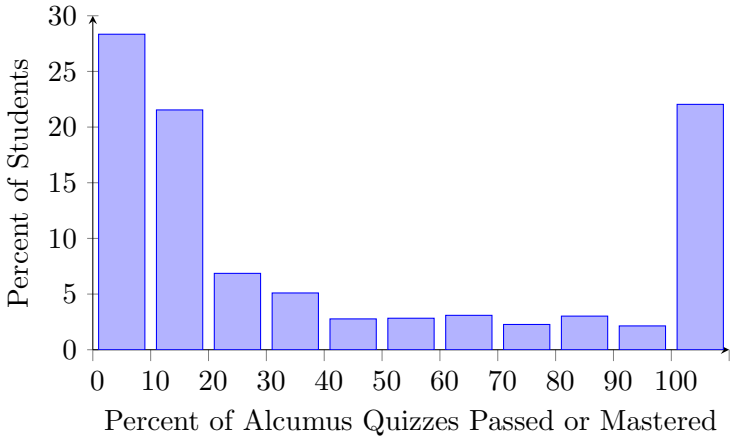


Figure 3.2: Student learning: percent of Alcumus quizzes passed

There is also less encouraging news in the figure. While the median student viewed 55% of the course components, the median student passed just 20% of the Alcumus quizzes. So the quiz-passing data look worse than the course-viewing data in the lower tail.

Even in the upper tail, the fraction who passed 80-100% of the quizzes is well below the fraction who viewed 80-100% of the course components: 27% vs. 37%. Even among students who viewed at least 80% of the course, there are many students who did not pass many Alcumus quizzes. Some of this phenomenon is due to students who appeared to just not bother with Alcumus—recall that the program was entirely voluntary and no school officials would ever see a student's results. But most of those who viewed 80% or more of the course without passing 80% of the Alcumus quizzes did put in substantial effort. The median number of Alcumus problems viewed by students in this group was 570. To make this concrete, two such students viewed exactly 570 Alcumus problems. One got 62.5% of the questions correct, passing 12 of the 36 quizzes. The other got just 45% of the questions correct, passing just 9. Each stuck with the course to the end: one clicked on 329 of the 330 non-Alcumus components and the other clicked on 328.

[6] The figure could also underrecord some students' engagement if, for example, they watched videos together with another student while logged into that student's account.

[7] Students must pass quizzes before they master them, so all students who have mastered have also passed.

Indeed, the course data contain many stories of remarkable grit among students who found the course very difficult. One student in our data never passed one particular quiz despite having given 145 correct answers on it: Alcumus never passed them because they took a total of 372 questions to obtain the 145 correct answers. The overall leader in Alcumus problems attempted was a student who viewed 8098 Alcumus problems. We infer that that student viewed so many problems because they aspired to master each of the quizzes despite the problems being hard for them—they got just 3693 of the 8098 problems correct. In the end, that student mastered 23 of the Alcumus quizzes and settled for passing the other 13.

On the one hand, Alcumus problems adapt to students' levels. The challenge problems, on the other hand, are both difficult and non-adaptive. They are, therefore, well-suited to provide an even more demanding examination of whether our top students learned the course material. The challenge problems again indicate that our top students are doing quite well. The left panel of Figure 3.3 shows percentage of viewed challenge problems that students who were in the top third of the engagement distribution got correct (on the first or a later attempt).[8] In the histogram we can see at the median such student correctly answered about 80% of the challenge problems they viewed. There is a left tail, however, with 20% of the students who are in the top third of engagement getting fewer than 40% of their viewed challenge problems correct.

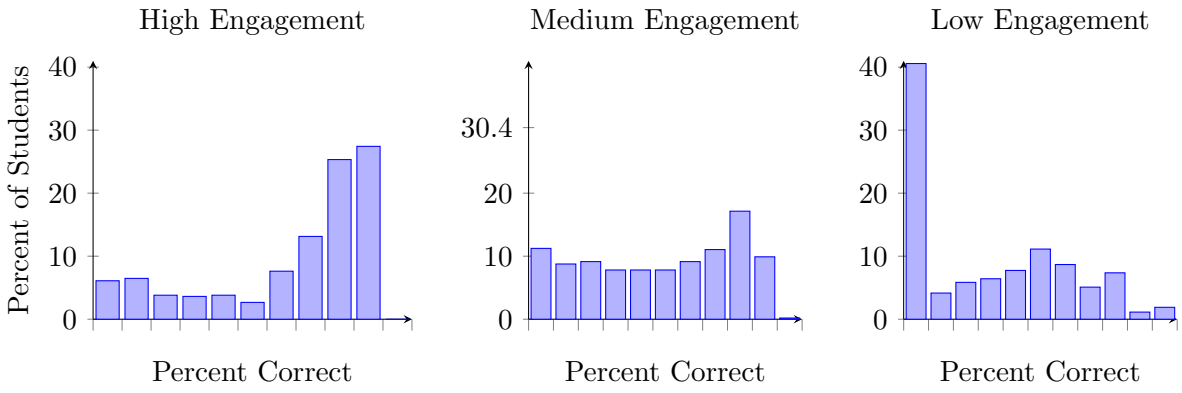


Figure 3.3: Student learning: fraction of challenge questions correct for students with high, medium, and low course engagement

The fact that the challenge problems are non-adaptive makes them well-suited for examining the relationship between student engagement and ability to solve problems. The middle and right panels of Figure 3.3 present similar histograms for students in the middle and bottom tercile of course engagement. Students in the middle panel clicked on 30-85% of the course components and those in the right panel clicked on fewer than 30%. The three histograms look quite different. Most importantly, the histogram on the right suggests that the difficulty of the course may be one factor that leads some students to make little progress: 40% of the students in this low-progress group

[8] These students clicked on at least 85% of the course components.

answered fewer than 10% of the challenge problems that they viewed correctly. Students with the middle level of engagement have nearly uniform heterogeneous experiences with the challenge problems: some get almost all correct, some get almost none, and many are in between.

Figure 3.4 presents a similar figure but reporting the fraction of Alcumus questions students answer correctly. It provides a nice illustration of the adaptive technology. Most students in all three terciles get about 70 percent of their Alcumus questions correct. (The two highest bars in the medium and low-engagment histograms are 60-70 and 70-80 percent correct. The two tallest bars in the high-engagement histogram are 70-80 and 80-90 percent correct.) Two other notable features are that the high-engagement tercile has an upper tail of very high performers who get 90+ percent correct and the low-engagement tercile has a set of low performers who get fewer than 10% correct. This could reflect that they are not trying, that they give up before Alcumus can adapt to their level, or that they are coming into the program too far behind to handle even its easiest problems.

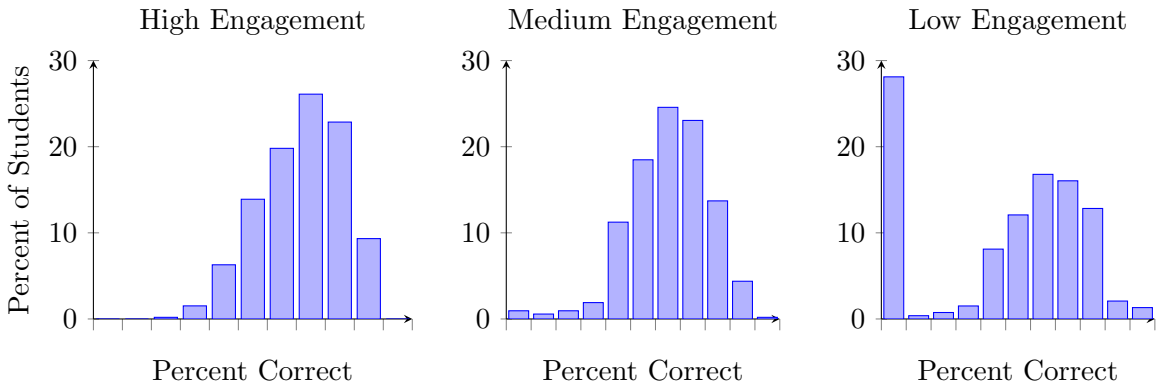


Figure 3.4: Student learning: fraction of Alcumus questions correct for students with high, medium, and low course engagement

As a single summary statistic of how well our selection and support procedures did at identifying talented students and helping them to successfully complete the AoPS course, 326 of our treatment students, or slightly over 20%, engaged with at least 80% of the course components, passed at least 90% of the Alcumus quizzes, and eventually solved at least 60% of the challenge problems that they viewed.

4. Program Design: Student Selection

While the primary goal of our experiment was to assess whether it was feasible to find students who might benefit from a multi-year intervention, a secondary goal was to gain insights on how such students might be identified. We obviously cannot yet assess whether students selected by any method were more likely to benefit from a long-run intervention, but we can use engagement and learning in the AoPS *Prealgebra 1* course as proxies.

In the US, the natural way to identify talented disadvantaged students would be with state standardized tests. This method has two primary concerns: the tests are typically not designed to distinguish among students at high performance levels; and many/most high-scoring middle-school students would have no interest in doing extracurricular math courses. In Tamil Nadu, we also had a third concern: many schools report implausibly high scores for their students on standardized tests. This concern motivated our decision to make in-person visits to 1441 schools to gather additional data on some of the school's high-scoring students. After these visits, we had four quantitative measures that we could use in the student selection: the 6th-grade end-of-year score on their standardized test (available in administrative data provided by the school department); a measure of students' engagement with the booklet we gave them, through both photos of their workbooks and engagement with an online portal to check answers; the score on a test we wrote assessing mastery of the booklet material; and a measure of "grit" obtained from other items on our test. Reflecting the concern that some schools were likely inflating standardized test scores more than others, we converted the students' standardized test scores to a within-school z-score.[9] Also, reflecting the fact that most students did cheat on our test as well, we replaced the scores of students that we identified to have cheated with the score one would have predicted given the student's standardized test score **if** the student scored higher than this on our test. We retained the actual test score if a student scored below the predicted score despite cheating.

As described earlier, we constructed eight different indexes that could be used to select students for the program and randomized at the school level which index we used to select students. The choices reflected a prior that the test score and booklet engagement measures would provide information about the likelihood of engaging with the AoPS course that would complement the information we obtain from the administrative data. The booklet-only, test-only, and grit-only selections were included primarily to broaden the set of student profiles for which we had observed outcomes.

Table 4.1 provides four measures of engagement/learning for the set of students selected by each method. The selection using only the administrative data fares worse than the other selections on all three metrics. The proportion of students selected purely on the basis of administrative data who met the "successful completion" benchmark[10] was just 12%. The selections using booklet engagement do well: the admin+booklet selection was the most successful under the first two metrics and the booklet-only selection was most successful under the third. In thinking about the success of the booklet-only, test-only, and grit-only selections relative to the admin-only selection it should be kept in mind that the booklet-only, etc. selections are making extensive use of administrative data:

[9] This conversion could be motivated by a model in which the true distribution of scores is the same in each school and a score-inflation process involves affine transformations that differ across schools.

[10] Recall we defined successful completion as clicking on at least 80% of the course components, passing 90% of the Alcumus quizzes, and solving at least 60% of the challenge problems that they viewed.

the first-stage selection of 9,709 students out of 73,533 was primarily based on the administrative data, and the description here is how we choose 2,490 students from the pool of 9,709 students with fairly high scores on standardized tests in the administrative data.

Selection Index	Treatment Schools	Engage-ment	Alcumus Passing	Successful Completion	Effect on Indx	
					Coeff.	Std.Err
Admin + Booklet + Test	118	0.59	0.43	0.21	0.13	(0.04)
Admin + Booklet	53	0.60	0.47	0.26	0.15	(0.04)
Admin + Test	127	0.51	0.37	0.18	0.04	(0.04)
Booklet + Test	49	0.54	0.40	0.21	0.12	(0.05)
Booklet only	49	0.58	0.45	0.29	0.09	(0.05)
Test only	50	0.58	0.44	0.23	0.10	(0.05)
Grit only	41	0.52	0.43	0.24	0.04	(0.05)
Admin only	77	0.45	0.33	0.12	—	—

Table 4.1: Statistics on engagement and learning for students selected by various methods

Our preregistered test for significance of the differences across the groups was the following: we regress an index of AoPS performance on the selection method—specifically, dummies for each dimension of the subtreatment assignment—and dummies for whether the student was male, disadvantaged, and attended an English-medium school, with standard errors clustered at the school level.[11] The right two columns of Table 4.1 present coefficient estimates and standard errors from such a regression with the admin-only selection as the omitted group. It indicates that several of the differences in effectiveness relative to the admin-only selection are significant. The admin+booklet selection has the largest point estimate, though most comparisons among other methods are not statistically significant.

The second way in which we examine selection is via regressions of measures of engagement/learning the student characteristics that were known at time when we selected students. As these methodologies selected students solely on the basis of right-hand-side variables in these regressions, they provide a valid way to investigate how one could have best predicted which students would be successful. Table 4.2 presents coefficient estimates from regressions with four different dependent variables. The first uses the same index of AoPS engagement/learning as above.[12] The others provide some insight into what is driving this result by separately examining effects on engagement, Alcumus quiz passing, and a dummy for "successful completion." The regressions include variables that could have been used to select students: the 6th grade administrative score, booklet engagement, our test, our grit measure, and dummy variables for whether the student's teacher

[11] The index places weight one-third on the fraction of items clicked, one-third on the fraction of problems answered correctly (with weight one half on challenge problems viewed one-half on Alcumus problems viewed), one-sixth on the fraction of items clicked by June 1, and one-sixth on the number of problems completed (with weight one half on challenge problems and weight one-half on Alcumus defined as quizzes passed).

[12] The AoPS index regression is the one that appears in our pre-analysis plan.

suggested the student as their top student or as another student who was not the top student but might respond well to a challenge. Each of the continuous variables are normalized to have standard deviation 1, so the coefficient estimates can be compared across measures as the effect of a one-standard deviation increase. The regressions also included several control variables not reported in the table: dummies for each dimension of subtreatment assignment, and dummies for male, disadvantaged, and English-medium students.

	Dependent Variable:			
Independent Variable	AoPS Index	Course Engagement	Alcumus Passing	Successful Completion
6th Grade Admin	0.050	0.047	0.042	0.010
	(0.01)	(0.01)	(0.02)	(0.02)
Booklet Engagement	0.092	0.092	0.082	0.068
	(0.01)	(0.01)	(0.01)	(0.01)
Test on Booklet	0.027	0.022	0.021	0.015
	(0.01)	(0.01)	(0.01)	(0.01)
Grit	0.013	0.011	0.016	0.014
	(0.01)	(0.01)	(0.01)	(0.01)
Teacher Topper	0.026	0.023	0.018	0.002
	(0.02)	(0.02)	(0.02)	(0.02)
Teacher Challenge	-0.032	-0.027	-0.038	-0.065
	(0.03)	(0.03)	(0.04)	(0.03)

Table 4.2: Statistics on engagement and learning for students selected by various methods

In the regression examining the index of engagement/learning, the Booklet Engagement variable has the largest coefficient and is highly statistically significant. We conclude that this method of identifying students who will benefit from an intervention using a smaller version of said intervention is supported in the data.

The 6th grade administrative test score and the test we gave on the booklet material are also statistically significant predictors of engagement/learning, suggesting one would also want to include them in an index used to select students, albeit with smaller weights. The grit and teacher-topper measures also have positive point estimates, but are not statistically significant. The non-top students that teachers suggest might benefit from a challenge are not as successful.

The estimates in the second column, focusing just on the fraction of course elements a student clicks on, are similar to the estimates in the first column. This is not surprising given that student engagement measure is a heavily weighted component of the index. The results on the fraction of Alcumus quizzes that students pass are also similar. The coefficient on the test now falls short of significance at the 5% level. In the column analyzing successful completion of the course only the booklet engagement measure is significant. The coefficient estimate indicates that a student who is one standard deviation above average in engagement with the booklet is 6.8 percentage points

more likely to successfully complete the course. This is quite large relative to the mean completion probability of 20%.

5. Program Design: Support and Incentives

Helping talent students to reach their potential requires much more than just providing them with access to online resources. All students in our treatment receive several types of support: we helped guide them through the AoPS sign up process, we provided them with 5G-connected tablets with which to do the courses, we sent periodic WhatsApp messages to them and to their parents encouraging them to work on the courses, we had ceremonies at the schools when they started the program, etc. We also randomize three treatment components to examine their incremental impact on student engagement and learning. We discuss results on subtreatment effectiveness in this section. One striking result is that tutoring support was not necessary for students to succeed in the online courses – in fact, students without tutors did substantially better!

5.1. Tutoring A number of previous papers have found that online education programs on their own may have limited impact. Tutoring would be costly and challenging to implement at scale, but our prior had been that it could be a very valuable complement and might be necessary to make online courses effective in our setting. As described earlier, we randomized our students into three treatment arms at the school level: 50% received no tutoring support, 37.5% were assigned to a group tutor, and 12.5% were assigned to a one-on-one tutor.

The top portion of Table 5.1 presents descriptive statistics on the engagement and learning of students randomized into the different tutoring treatments. Whether one looks at the number of course elements students clicked on, the number of Alcumus quizzes passed, or a successful completion dummy, the students without tutors did substantially better than the students in either tutoring arm. The results are most stark for successful completion: it is 30% for students without tutors, 18% for students with one-one-one tutors and 11% for students with group tutors. The significance test we preregistered for this portion of analysis was to regress our AoPS index on dummies for each of the subtreatments.[13] The right two columns show coefficient estimates and standard errors from this regression. They indicate that the relative non-success of the tutoring treatments is highly statistically significant. From a practical perspective this finding can be seen as very good news – the program is much cheaper and easier to scale up if tutors are not needed, and they very much appear not to be necessary.

Why are we not finding tutoring effects here when such support has been found to be valuable in other settings? It could be that the received wisdom about the value of tutoring support is incorrect. But it seems more likely that there is something special about our environment and/or

[13] As before these regressions also included dummies for male, disadvantaged, and English-medium students.

our intervention that makes tutoring less valuable. One obvious possibility would be that our tutors simply were not good. We do not think that this is what is going on. After not seeing tutoring benefits in an earlier pilot, we put a great deal of effort into tutor recruitment, training, and monitoring for this RCT. Our tutors were students from good technical universities and our impression was that most tutors were enthusiastic about the program, worked diligently, and taught well. One feature of our environment that could play a role is that most students did not successfully complete the course. Students in group tutoring in particular, might have made slow progress because they learned from others that they were not supposed to work hard or would have found it awkward to be ahead of the group. Another feature our environment that could play a role is that AoPS courses may engender a beneficial online addiction, and it could be that time spent with the tutor takes time away from watching AoPS videos and doing Alcumus, and makes students less likely to get hooked. Future years of the program may provide opportunities to help disentangle the multiple potential explanations.

Subtreatment	Number of Schools	Engage-ment	Alcumus Passing	Successful Completion	Effect on Indx	
					Coeff.	Std.Err
Individual Tutoring	68	0.49	0.37	0.18	-0.16	(0.02)
Group Tutoring	215	0.47	0.32	0.11	-0.15	(0.04)
No Tutoring	281	0.64	0.50	0.30	—	—
Hi Tech Lab Intent	156	0.59	0.45	0.22	0.04	(0.03)
Hi Tech Lab Control	408	0.54	0.40	0.21	—	—
Enforced Progress Req.	188	0.56	0.42	0.20	0.05	(0.03)
Unenforced Progress Req.	184	0.59	0.45	0.25	0.03	(0.05)
No Progress Req.	192	0.51	0.38	0.19	—	—

Table 5.1: Statistics on engagement and learning for students in various subtreatments

5.2. Hi Tech Labs The government of Tamil Nadu is in the midst an ambitious project to construct Hi Tech Labs in government schools. Our Hi Tech Lab subtreatment, randomly assigned to 25% of schools, involved contacting school prinicpals and asking them if students could work on AoPS side by side in the school's Hi Tech Lab at some time (often afterschool). Estimates of the effect of the program should be thought of as intent-to-treat estimates because it was sometimes not possible to do this. The middle panel of Table 5.1 presents descriptive statistics. The students chosen for the Hi Tech Lab subtreatment accomplished slightly more on each of the dimensions shown, but the difference in the AoPS index is not statistically significant.

While the HTL effect is not significant, we continue to believe that the HTL version of the program is superior and the benefit fails to be apparent only because of unfortunate timing. We launched the program itself in early February, but with all of the other things going on were only able to start arranging with schools to offer Hi Tech Lab access in early March. Students therefore only got one month of exposure to the Hi Tech Lab treatment before the Lab access was paused for final

exams and then unavailable in the summer vacation (which runs from mid-April to early June). Figure 5.1 graphs our student engagement measure (the number of items clicked on) over time for the HTL treatment and control groups. The groups were making quite similar progress up to the two-week period in which we started to roll out the subtreatment. The HTL group then makes progress at approximately twice the rate of the control group over the next month. But lab access is then interrupted by final exams and summer vacation (which together run from early April to early June). The advantage of the HTL group fades over this period. This graph leads us to conjecture that the HTL treatment would have been more effective if the course had been offered entirely during the school year with uninterrupted lab access.

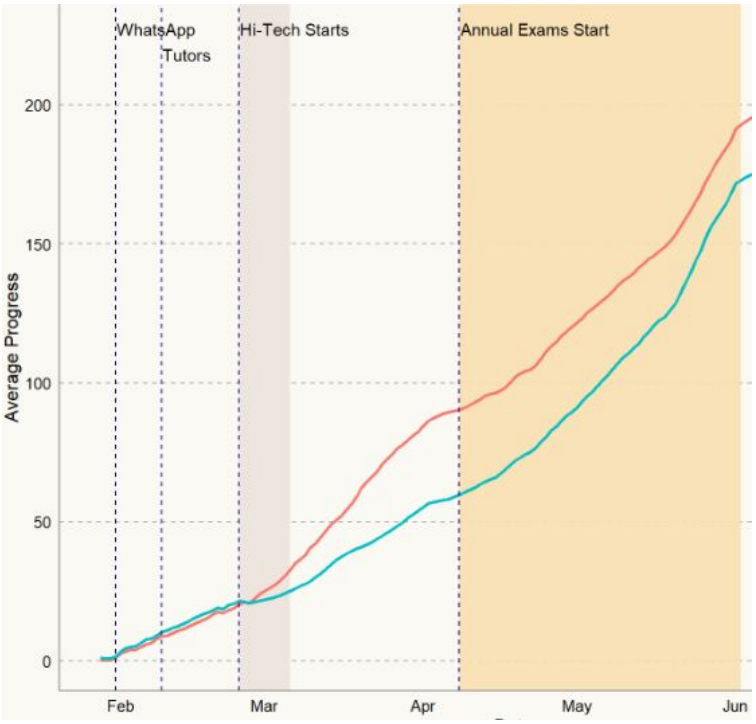


Figure 5.1: Progress of the Hi Tech Lab treatment and control groups over time

5.3. Progress Requirements A third dimension on which we varied the program was in the treatment of students who were not making sufficient progress in the course. Two subtreatment groups received messages warning them that tablets would be taken away if they were not making sufficient progress. A third did not receive any messages. The groups that received the warning messages was a bit more engaged then the group that did not, but the differences are not very large and again fall short of statistical significance.

In part, the failure of the warning messages to be effective may reflect that our messages did not become highly specific until late in the summer vacation. (Taking away tablets during the summer vacation would have incurred substantial costs so we wanted to avoid doing that.) No tablets were

actually taken away until the final month of the program. Even if the incentive effects are small, we think it is advisable to include in the program. The data give no reason to think that the incentive effect is negative, and if students are using their tablets to watch YouTube video rather than to work on AoPS we would conjecture that it is better for them to have the tablet taken away.

6. Program Effectiveness: Heterogeneity

In the previous discussions we have noted in a few places that regressions included controls for gender, disadvantaged status, and whether students studied in English or Tamil. In this section, we present some results on differential engagement and performance.

One interesting set of observations concern gender. There is a gender gap in college STEM majors in many countries. The gender gap in elite Indian education is particularly large. It has been reported, for example, that in the mid 2010's only 8% of students at the Indian Institutes of Technology were female.[14] In light of this gender gap it is striking that two thirds of the students selected for our program are female. (Recall that our selection was gender blind and relied solely on administrative scores, booklet engagement, and assessments of knowledge and grit from the test we gave when visiting schools. Moreover, the female students in our courses performed better, both in terms of course engagement and in terms of passing Alcumus quizzes. They also had an advantage on answering questions correctly, but this gap was a little smaller.

About half of our students are in English-medium schools and the other half take most of their courses in Tamil. One concern that both we and the Tamil Nadu education department had about our program is that the fact that the AoPS course is in English might make it inappropriate for Tamil-medium students.[15] An encouraging result in this regard is that there were no differences in course engagement between the English- and Tamil-medium students. Tamil-medium students did, however, pass fewer Alcumus quizzes and have substantially lower percentages of correct answers on the challenge problems. This could be due either to Tamil-medium students have not learned the material as well or to their misreading some questions.

About 24% of the students selected for participation are members of a scheduled caste or tribe. These students made somewhat less progress on the course and are only 17% of our successful completion group.

[14] See "IIT Madras offers Unparalleled Academic Flexibility to Students," IIT Madras Press Release, October 19, 2020. IIT Madras has since started an affirmative action program for female students and now has about 20% female undergraduates.

[15] We added Tamil subtitles to the course videos, but the videos themselves, the written lessons, and all problems were only in English.

7. Program Effectiveness: Endline Exam

We have primarily relied on student performance on Alcumus quizzes and challenge problems as evidence that students are learning from the AoPS course. In future drafts of this paper we will also examine two other measures of student learning. These alternate measures can be compared between treatment and control students.

One measure is performance on administrative exams at the end of 7th grade (two months into the course) and early in 8th grade. We anticipate having these data, but have not yet received them. Limitations of this data include that the tests are not designed to distinguish among high achievers and that some schools seem to inflate their students scores. We will probably not have these data for a few months.

A second measure is performance on an endline exam that we conducted in July (when we revisited schools to reclaim tablets). Rather than testing knowledge of the AoPS course, the exam we gave was intended to assess students' abilities to solve highly challenging problems on topics that are included in the Tamil Nadu grade 7 or 8 syllabus.

The exam was even more challenging than we had intended, with the median control student only getting about 1.2 of the challenging questions correct. Although we sent proctors to the schools with instructions to prevent students from cheating and designed a very cheating resistant test, there were again some problems with students copying off each other.[16] Each of these factors limits our ability to draw strong conclusions. Preliminary estimates indicate that treatment students answered 20% more of the challenging problems. But this conclusion is highly tentative as we are still making sure we are dealing adequately with the cheating.

8. Conclusion

The RCT described in this paper is the first stage of a larger project intended to provide assistance to talented students from disadvantaged backgrounds in Tamil Nadu over a several year period. The idea behind the larger intervention was to use engagement with and performance in one online class as a way to identify students who would take advantage of a longer-run intervention that will begin with other similar perspectives. From that perspective we see the outcome as a great success: we have identified a number of students who engaged with and did very well in a challenging course; and we identified a number of additional students who displayed a great deal of grit while struggling to adapt to the level of the course.

Our results on methods for selecting students indicate that a simple measure of student engagement

[16] To deter cheating we had 10 distinct test forms. The test forms differed both in the order in which questions appeared and in the numbers that were used in some questions. The tests were not multiple choice and most students were getting most questions wrong, so it was easy to see when students had copied of each other.

with one chapter of a physical textbook was a valuable predictor of whether students would engage with the online course. We hope that this is both useful finding for others planning similar interventions and an early indicator that the students we identified as doing well in one online course will do well in future online courses.

A striking finding on program design was that it did not seem necessary to provide students with tutors to make it possible for them to succeed. Our simple WhatsApp message groups seemed to be enough. And the group tutors particularly seemed to in fact have a negative effect on student progress. It could be that things would work out differently in future stages of our project – it seems plausible that a group tutoring meeting where students learn that other students in the group have all made a lot of progress has a very different effect than a group meeting in which students learn that not making much progress is the norm.

In this experiment we have mostly relied up students passing Alcumus quizzes and solving challenge problems as evidence that students are gaining from the AoPS courses. In the long run, we will be more concerned with whether students are moving forward on the path to STEM careers. The schools from which we are drawing students have essentially no students who ever reach the top levels of the Indian educational hierarchy, so it will be interesting to see where our program can take our study participants.

References

Agarwal, Ruchir, Ina Ganguli, Patrick Gaulé, and Geoff Smith (2023) "Why U.S. Immigration Matters for the Global Advancement of Science," *Research Policy*, 52 (1), 104659, 10.1016/j.respol.2022.104659.

Agarwal, Ruchir and Patrick Gaule (2020) "Invisible Geniuses: Could the Knowledge Frontier Advance Faster?" *American Economic Review: Insights*, 2 (4), 409–424, 10.1257/aeri.20190457.

Agarwal, Ruchir, Patrick Gaulé, and Yuyan Jiang (2024) "Finding Young Einsteins: Olympiads and STEM Talent Discovery," *Unpublished*.

Ahuja, Amrita, Douglas B Marshall, and Fernanda Ramirez-Espinoza (2024) "The Journey of Mathematically Gifted Students in Chile from Primary School to Higher Education."

Banerjee, Abhijit V., Swati Bhattacharjee, Raghabendra Chattopadhyay, Esther Duflo, Alejandro J. Ganimian, Kailash Rajah, and Elizabeth S. Spelke (2025) "Children's Arithmetic Skills Do Not Transfer between Applied and Academic Mathematics," *Nature*, 10.1038/s41586-s̄024-s̄08502-s̄w.

Carlana, Michela and Eliana La Ferrara (2021) "Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic," *HKS Faculty Research Working Paper Series*, RWP21-001.

Cohodes, Sarah R, Helen Ho, and Silvia C Robles (2024) "Diversifying the STEM Pipeline: Evidence from STEM Summer Programs for Underrepresented Youth."

Ellison, Glenn and Ashley Swanson (2016) "Do Schools Matter for High Math Achievement? Evidence from the American Mathematics Competitions," *American Economic Review*, 106 (6), 1244–1277, 10.1257/aer.20140308.

Koedinger, Kenneth R, John R Anderson, William H Hadley, and Mary A Mark (1997) "Intelligent Tutoring Goes To School in the Big City," *Carnegie Mellon University*, 10.1184/R1/6470153.v1.

Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian (2019) "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India," *American Economic Review*, 109 (4), 1426–1460, 10.1257/aer.20171112.

Steegh, Anneke M., Tim N. Höffler, Melanie M. Keller, and Ilka Parchmann (2019) "Gender Differences in Mathematics and Science Competitions: A Systematic Review," *Journal of Research in Science Teaching*, 56 (10), 1431–1460, 10.1002/tea.21580.

Ugarov, Alexander (2019) "Talent Misallocation across Countries: Evidence from Educational Achievement Tests," *Unpublished*.

A. Methodology for initial screen

Here, we describe the methodology used to select the 9,709 students who were eligible for selection into our RCT.

1. Beginning with our population of 112,630 students in 3,494 schools, we first identified students we believed had a high probability of succeeding in the program. At each school, we identified the number of students that meet the following criteria:

   - A "very high scorer" is a student with a raw score of 90 on their 6th grade final exam scores, and a within-school z-score of over 2.
   - A "high scorer" is a student with a raw score of 80 on their 6th grade final exam scores, and a within-school z-score of over 1.5.

   We selected a school for inclusion if either

   - the school contains at least 1 very high achiever, or
   - the school contains at least 2 high achiever.

   This procedure allows us to target selection towards schools where we expect to find students that can benefit from our program, and selected $1,355$ schools that enroll $73,533$ 7th graders in total.

2. At included schools, our research team selected the following students were offered the sample AoPS content:

   - Students with a raw score of over 70, a z-score of over 1.2, and with $30z + \text{score} > 125$. (All criteria needed to be met to be chosen.) In doing so, we attempted to capture students with hgih achievement in mathematics..
   - For schools with more than 5 students selected using the above criteria, we also, with probability $\frac{1}{2}$, selected $\sim 15\%$ more students randomly from those above the 75th percentile within their school. This allowed us to explore how selection would work for students lower in the achievement distribution.

   This procedure selected $7,272$ students across the $1,355$ schools.

3. As teachers may have more informative signals about student ability than that observed in the adminsitrative data, the final set of students was chosen using the following procedure.

   (a) At each school, we described the study and sample content to the math teacher. We then ask them to identify the top 5% of students in mathematics in their school (minimum of 1) to receive the sample AoPS content.

   (b) We then revealed the list of students described in step (2).

   (c) Finally, we asked the teacher to identify the top 2.5% of students (minimum 1) who have "high potential" – students that may benefit from our program, even though their test scores may not necessarily reflect it.

   This resulted in our final sample of $9,709$ students.