Which Students Benefit from Independent Practice? Experimental Evidence from a Math Software in Private Schools in India*

Andreas de Barros†
Alejandro Ganimian‡
Anuja Venkatachalam§

October 27, 2020

Abstract

This study is one of the first evaluations of independent (i.e., self-guided) practice in math in a developing country. We randomly assigned 4,461 students in grades 4-7 in high-performing private schools across seven Indian cities who were using a computer-assisted learning software to: (a) a control group, in which they moved from one unit to the next upon completion; or (b) a treatment group, in which they had to complete practice exercises before progressing to the next topic. After six months, the additional practice had a precisely estimated null effect on the math achievement of the average student. However, treatment students with low initial performance outperformed their control counterparts by 0.14 standard deviations (SDs). Our results suggest that independent practice may help private-school students in need of catching up. This is an important finding given the large and growing enrollment in this sector. Yet, it also suggests that, if affluent private schools provide more opportunities for practice, this may exacerbate inequities of the broader school system.

Keywords: computer-aided learning, India, math instruction, practice exercises, private schools

JEL classification: C93, I21, I22, I25

Study pre-registration: Pre-registered at the AEA RCT Trial Registry (AEARCTR-0002455)

*We gratefully acknowledge the funding provided by the Douglas B. Marshall, Jr. Family Foundation for this project. We especially thank Karthik Muralidharan for his involvement as a collaborator in the early stages of this project and for subsequent discussions. We thank Pranav Kothari, Aarthi Muralidharan, Sridhar Rajagopalan, Maulik Shah, Nishchal Shukla, and Gayatri Vaidya for making this study possible. The usual disclaimers apply. The authors have no conflicting interests to declare.

†Postdoctoral Associate, Department of Economics, MIT. E-mail: debarros@mit.edu.

‡Assistant Professor of Applied Psychology and Economics, New York University Steinhardt School of Culture, Education, and Human Development. E-mail: alejandro.ganimian@nyu.edu

§Data Journalist, DataLEADS. E-mail: anujav04@gmail.com

There is broad consensus among math educators about the importance of practice for primary- and middle-school students (see Woodward et al. 2012). Yet, the type of practice that has been found effective is often preceded by a considerable degree of scaffolding from teachers, either through "task lists" of specific steps (see, e.g., Hohn and Frey 2002; Verschaffel et al. 1999), "self-questioning" checklists (e.g., Jitendra et al. 1998, 2011) or processes (Cardelle-Elawar 1990, 1995; King 1991; Kramarski and Mevarech 2003; Mevarech and Kramarski 2003). Far less is known about self-guided practice that is not necessarily matched to specific "metacognitive" strategies that provide students with a structure to think about how to approach a type of problem. This type of practice is becoming increasingly important with the advent of coronavirus, which has suddenly required that students take a more independent role in their education.

More specifically, we still know far too little about whether (and if so, how) independent practice affects students at varying levels of preparation for schooling differently. Most of the evaluations of practice-based interventions cited above do not estimate heterogeneous effects by baseline achievement, and many of those that do lack sufficient statistical power to detect them, so it has been challenging to make sense of null interaction effects in this literature. Ex ante, it is not clear what to expect. On the one hand, low-performing children may benefit from learning at their own pace and practice exercises may encourage them to do so. Accordingly, multiple impact evaluations of software-enabled practice in these contexts have consistently found small-to-moderate effects on student achievement (see, e.g., Bettinger et al. 2020; Lai et al. 2015, 2013, 2016; Mo et al. 2015; Pitchford 2015). On the other hand, these children are often several grade levels behind curricular expectations, so additional practice may not be as helpful as remedial work that focuses on building foundational skills and remedying misconceptions. That may be the reason why software that dynamically adjusts to the performance of each student has yielded even larger effects on achievement than those that review the material covered in school during a given week (e.g., Banerjee et al. 2007a; Muralidharan et al. 2019).

This paper presents one of the first studies of the effects of technology-enabled independent practice in a developing country. During the 2017-2018 school year, we partnered with an educational-assessment firm in India to randomly assign 4,461 students in grades 4-7 in private schools across seven Indian cities that were using a computer-assisted learning (CAL) software to: (a) a control group, in which students complete a set of "learning exercises" (which present new concepts) on a given topic in math (e.g., operations with fractions) and then move on to the next topic; or (b) a treatment group, in which students also complete the learning exercises but are then asked to complete a set of "practice exercises" (which seek to build procedural knowledge) before moving on to the next topic. Students could interact with the software during or after school, but the bulk of usage occurred at home (66% on average, on any given day during the study), which makes it particularly relevant for the current context of the pandemic. We can verify that the intervention was implemented mostly as intended: during the six months of the study, the median student across both experimental groups interacted with the CAL software for more than 700 minutes and treatment students interacted with the practice exercises for more than 60 minutes. The three topics in which students completed the most amount of practice exercises were: measurement (14% of all exercises), fractions (10%), and number theory (9.1%). Further, interaction with the software remained relatively constant during the study—typically, students used the software between 20 and 40 minutes per week. Treatment students collectively completed nearly half a million practice exercises.

We report two main sets of results. First, after six months, we find that practice exercises had a precisely estimated null effect on math achievement, as measured by an independent test designed by the research team (not by the software developers): the average treatment student performed 0.014 standard deviations (SDs) better than their counterpart in the control group, and the difference is statistically insignificant. We can rule out positive effects larger than 0.062 standard deviations (as per the upper limit of the estimate's 95% confidence interval). In fact, we observe effects consistently estimated around zero across all topics and skills in the assessment. Further, we do not see any relationship between the number of practice exercises completed by the average student and math achievement.

Yet, these average impacts mask non-trivial heterogeneous effects. Initially low-performing students (i.e., those in the bottom quartile of baseline achievement within their grade level) outperformed their peers in the control group by 0.136 SDs ($p < 0.01$). In fact, they improved in two of the three content domains (numbers and geometry) and in all three cognitive domains (knowing, applying, and reasoning) assessed in the endline test. We examine whether low performers who spend more time on the software and do more practice exercises fare better, but to our surprise, we do not find that this is the case. We do not find any statistically significant heterogeneous effects by students' school, grade, or sex.

This study makes several key contributions to research on student learning in developing countries. First, it draws attention to the issue of heterogeneity in the benefits of independent practice. Prior studies in math pedagogy have sought to recruit students from disadvantaged backgrounds to understand the impact of independent practice on a segment of the student population who stands to benefit from additional exposure to the material. Yet, for the most part, this literature relies on small samples, which prevent researchers from comparing the effect of independent practice across different types of students. The sudden disruption imposed by coronavirus on school systems make this question particularly timely.

Second, this study also contributes to the growing literature on private schools in India. Prior studies have established that these schools, which serve one in three school-aged children in the country (Kingdon 2020), outperform public schools (Muralidharan and Kremer 2008), their advantage is mostly due to student selection rather than better instruction (Singh 2015), but their lower teacher salaries make them more efficient (Muralidharan and Sundararaman 2015). Our study adds to this growing literature, both documenting business-as-usual progress for private-school students across major Indian cities and identifying an intervention that may advance the achievement of this large (and growing) segment of the Indian school system.

More generally, our study illustrates the potential of leveraging CAL software for rapid-cycle randomized controlled trials that can shed light on the pedagogical approaches that work best. Often, decisions about software features are made by education specialists working alongside developers, informed by research based on small, convenience samples in the U.S. In a few cases, it is also guided by so-called "A/B testing", in which a specific feature of the software is offered to a random subset of its users. Yet, as our paper demonstrates, partnerships between researchers and software developers can produce studies that address questions of broader interest to the field of education, ensure that minimal conditions for causal inference are met (e.g., equivalence at baseline and post-attrition), and that the results are shared (and scrutinized) by the relevant scientific community.

The rest of the paper is structured as follows. Section 1 presents the context, study design, and intervention. Section 2 describes the data. Section 3 discusses the empirical strategy. Section 4 reports the results. Section 5 discusses implications for research and policy.

1 Experiment

1.1 Context

Schooling in India is compulsory and free from ages 6 to 14 (Ministry of Law and Justice 2009). Primary education runs from grades 1 to 5 and upper primary runs from grades 6 to 8; grades 1 to 8 are collectively referred to as "elementary education". The Indian school system included 840,241 primary schools, 287,265 upper-primary schools, and 48,543 primary schools with secondary grades in the 2016-2017 school year (NIEPA 2018c). Private-unaided schools serve nearly a third of students in elementary grades (58,364,364 students, or 31% of total enrollment).

We conducted this study in partnership with Educational Initiatives (EI), a leading assessment firm in the country that developed the CAL software that we used to randomly assign students to practice exercises (described in greater detail in the Intervention sub-section). We established this partnership as a multi-year project to leverage both the vast item bank of the CAL software in math and other subjects and its high degree of penetration across the country to use randomized experiments to answer questions of import to educators. The partnership, dubbed the Learning Lab, was led by Karthik Muralidharan at the University of California, San Diego and Sridhar Rajagopalan at EI and funded by the Douglas B. Marshall Foundation. We were co-principal investigators and research associates on this project.

We conducted our study in high-performing private schools across seven Indian cities: Ahmedabad and Rajkot (in the state of Gujarat), Faridabad (Haryana), Ghaziabad (Uttar Pradesh), Kolkata (West Bengal), New Delhi (Delhi), and Tiruchirappalli (Tamil Nadu). As Table A1 in Appendix A shows, in nearly all of these cities (except Kolkata), private schools account for the majority of student enrollment in primary education and, in most cases, they also outnumber public schools. Yet, there is considerable variation in the share of students enrolled in private schools, from 18% in Kolkata to 77% in Faridabad. There is also some variation in internal-efficiency (i.e., repetition and dropout) rates, but "no detention" policies and differences in data collection and reporting procedures across states make it difficult to interpret these figures.[1]

These cities also vary widely in their learning outcomes. As Table A2 shows, students in grade 5—the only grade in our study covered by the National Achievement Survey or NAS—performed very different on the latest national assessments of math and language. (Importantly, however, only government and government-aided schools participate in NAS). In some cities (e.g., Rajkot) most students master curricular expectations; in others (e.g., Tiruchirappalli), only a minority does so (see also ASER 2019; de Barros and Ganimian 2020). It is possible that the public-school students that participate in NAS perform differently from private-school students (see Muralidharan and Kremer 2008), but we do not have comparable achievement data for the latter. Reassuringly, however, one longitudinal study of public and private schools in an Indian state found that public and private schools add similar value in urban areas (Singh 2015), suggesting that differences between public and private schools may be less of a concern in the cities that we study.

1.2 Sample

The sample for the study included 4,461 students from grades 4 to 7 across nine private schools in the aforementioned cities. We drew a convenience sample of schools because this was the first study of the Learning Lab (see description of the initiative in the Context section) and we needed a group of schools that met all of the hardware and software requirements to deliver the CAL software during the study period without major disruptions. We invited 12 private schools to participate. Nine of those schools agreed to participate. We sought informed consent from principals and teachers at those schools.

Attrition from the study was low: 4,001 of the 4,461 students who took the baseline assessment (90%) also took the endline assessment. We found no differential attrition by experimental group:

[1] At the time of our study, all cities in our sample had a "no detention" policy, so repetition rates may only reflect exceptional cases. In 2019, an amendment to the Right to Education (RtE) Act of 2009 ended this policy.

11% of control students and 9.4% of treatment students attrited, but the difference is not statistically significant.

1.3 Randomization

We randomly assigned the 4,461 students in our sample to: (a) a control group, in which they first completed a set of "learning exercises" (which presented new concepts) on a given topic in math (e.g., operations with fractions) and then moved on to the next topic; or (b) a treatment group, in which students also completed the learning exercises but were then required to complete a set of "practice exercises" (which seek to build procedural knowledge and fluency) before moving on to the next topic.

We stratified the randomization by grade and whether students performed below or above the median in their class in the CAL platform exercises completed prior to the study. Within each stratum, we randomized students individually. This approach maximizes statistical power, but its main drawback is that it allows for spillovers across students in the same classroom. Given that the benefit of the intervention stems from individual practice, however, we think this is unlikely to be a major concern.

Control and treatment students were comparable on their baseline achievement and sex, regardless of whether we compare all students present at baseline or only those who also took the endline assessment (i.e., non-attritors, see Table 1). In fact, not just the means, but the distribution of baseline achievement was quite similar across experimental groups (see Figure A1 in Appendix A).

1.4 Intervention

All students were using a CAL software called "Mindspark" at the start of the study.[2] The software was developed by Educational Initiatives (EI), a leading assessment firm in India, over a 10-year period. It has been used by over 500,000 students, it has a database of over 45,000 questions, and it administers over 2 million questions across its users every day. It can be delivered during the school day, before or after school at stand-alone centers, and through a self-guided online platform. The after-school version was recently evaluated through a randomized experiment and found to vastly improve the math and reading achievement of primary and middle-school students in Delhi (Muralidharan et al. 2019). In our study, students had access to the in-school version, which is

[2]We do not know the date in which each school started using the software, but we can estimate it using the number of practice exercises that students had completed by the start of the study and the time it took students to complete each exercise during the study. At baseline, the average student had attempted 843 practice exercises (the median student had attempted 683). Treatment students completed 230 such exercises during the study or 9.2 per week. This suggests students had been exposed to the software for 92 weeks prior to the study.

currently used by more than 100,000 students in 300 private schools in India and abroad (including some Arab states in the Persian Gulf).

There are two important differences between the private-school version of software (which we use in this study) and the after-school version (which was previously evaluated) that are worth highlighting to interpret our impact estimates. First, the private-school version mostly presents learning and practice exercises at or above grade level: across treatment students, 58% of all practice exercises they completed during the study were at grade level, and 20% were *above* their enrolled grade level (see Table A3). The after-school version caters to children from disadvantaged schools, and consequently emphasizes material below grade level. Second, the topics and skills covered by the private-school version is determined by each students' teacher, based on the curriculum and any additional practice he/she deems appropriate. The content and cognitive domains of the material presented in the after-school version is largely determined by a diagnostic test, which students take when they first interact with the software. Therefore, our impact estimates do not confound the effect of practice with that of the dynamic adaptation of the material, which is largely shut down in the private-school version of the software.[3]

We are not interested in evaluating the impact of the software; instead, we use it to randomly assign students to different levels of independent practice in math. Specifically, control and treatment interacted with the software in a similar manner (Figure 1). First, students are prompted by the software to select a math topic (e.g., fractions); each topic includes between eight and ten units (e.g., subtraction of fractions). Then, they complete a set of learning exercises that seek to introduce that unit (on average, students complete 16 fill-in-the-blank and multiple-choice questions per unit). Next, the experience of control and treatment students differ: students in the control group move on to the next unit after completing the learning exercises, whereas those in the treatment group are required to complete practice exercises to consolidate their understanding of the concepts and procedures taught through the learning exercises.[4]

The learning exercises that all students complete differ from the practice exercises that only treatment students are required to complete in four ways. First, as mentioned above, learning exercises introduce new concepts and procedures, whereas practice exercises focus on helping students develop their procedural knowledge (i.e., knowledge about the algorithms to be followed to solve a specific problem) and fluency (i.e., capacity to solve problems rapidly). Second, learning exercises are untimed, but students are given eight to ten minutes (depending on the topic) to complete each practice exercise. Students who are unable to complete a practice exercise in the

[3] To clarify, there is some degree of dynamic adaptation, but within a much more narrow set of grades than in the after-school version evaluated by Muralidharan et al. (2019).

[4] Importantly, control students previously had access to the practice exercises; their access was temporarily deactivated during the study. Given that we administered the baseline instruments on slightly different dates (see the Data section), some control students accessed a few exercises between the baseline and the deactivation: the mean control student spent 3.4 minutes interacting with practice exercises during this period, and the median control student only 1.7 minutes, confirming that their usage of practice exercises during the study was minimal.

allotted time are allowed to attempt it again (with a reset timer) during their next session (to avoid delaying students' progression from one unit to the next). Third, learning exercises are the same for students at all levels of initial achievement, but practice exercises are categorized in three difficulty levels (low, medium, or high), which are presented sequentially (students who complete low-difficulty exercises graduate to medium-difficulty exercises, and so on, regardless of whether the exercises are answered correctly).[5] Finally, while all units include learning exercises, not all of them include practice exercises.[6]

2 Data

We collected two main types of data: (a) students' achievement, before and after the intervention, to check for baseline equivalence and estimate impact; (b) students' usage of the CAL software and interaction with the intervention, to verify implementation fidelity and estimate the relationship between the number of practice exercises completed and achievement; and (c) information on students' grade and sex (we did not conduct a student survey).

2.1 Student achievement

We administered student assessments of math learning at baseline (before the intervention) and endline (six months after the start of the intervention).[7] These assessments evaluated what students ought to know and be able to do according to international standards, including three content domains (numbers, geometric shapes, and measurement) and three cognitive domains (knowing, applying, and reasoning). The distribution of items across content and cognitive domains was based on the assessment framework of the 2019 Trends in International Math and Science Study (TIMSS) for grade 4 (Mullis and Martin 2017).

Each test had 35 multiple-choice items. We drew on items from international assessments (e.g., TIMSS, PISA, Young Lives), domestic assessments (e.g., Quality Education Study, Student Learning Survey), and previous impact evaluations in India (e.g., the Andhra Pradesh Randomized Studies in Education or APRESt). We included items from a wide range of difficulty to reduce the possibility of "floor" effects (i.e., students not answering any questions correctly) and "ceiling" effects (i.e.,

[5] Unfortunately, we do not have data on learning exercises; only on practice exercises (for treatment students).

[6] We cannot calculate the share of units with practice exercises from our data, but we can calculate the percentage of all days in which students interacted with the CAL platform in which they also completed practice exercises. The average treatment student saw at least one practice exercise in 75% of all days in which he/she used the platform. Put differently, three of each four days that a treatment student used the software involved practice exercises.

[7] Different schools conducted the baseline assessment and started using the software on slightly different dates (see Table A4). However, throughout this paper, we limit our analysis to a common six-month period, starting on September 11, 2017 and ending on March 11, 2018.

students answering all questions correctly). At baseline, we had different assessments for each grade (i.e., grades 4 to 7), with four versions of each test to prevent cheating among students. At endline, we consolidated the assessments and had one for grades 4-5 and another one for grades 6-7, with two versions of each test.[8]

We used a non-equivalent anchor test (NEAT) design to link results across administrations (for a discussion of this design, see Kolen and Brennan 2004). We included an "anchor test" with overlapping items across rounds of data collection and we scaled the results for both rounds concurrently using a two-parameter logistic Item Response Theory (IRT) model.

2.2 Students' time on CAL platform and responses to exercises

We also obtained access to data on students' time interacting with the CAL platform and with the learning and practice exercises, as well as on whether their responses to these exercises were correct, to verify that the trial and the intervention were implemented as intended and to estimate the dose-response relationship between practice and achievement. Each student was assigned a unique login and password, which allowed us to track his/her usage and responses during the study. Using each student's Internet Protocol (IP) address and time of login, we can also determine whether he/she is using the software in school or at home.

3 Empirical strategy

We estimate the effect of the offer of practice exercises (i.e., the intent-to-treat or ITT effect) by fitting the following model:

$$Y^t_{igs} = \alpha_{r(gs)} + \beta T_{igs} + \theta Y^{(t-1)}_{igs} + \epsilon^t_{igs} \tag{1}$$

where Y^t_{igs} is the math achievement of student i in grade g and school s at time t (endline), $r(gs)$ is the randomization stratum of grade g and school s and $\alpha_{r(gs)}$ is a stratum fixed effect, T_{igs} is an indicator variable for random assignment to treatment, and $Y^{(t-1)}_{igs}$ is math achievement at time $t-1$ (baseline). The parameter of interest is β, which captures the causal effect of the intervention. We fit variations of this model that interact the treatment dummy with students' grade, sex, and baseline achievement (continuous or by within-grade quartile) to understand whether the intervention is more helpful for some sub-groups of students.

[8] The baseline tests can be accessed at: https://bit.ly/2MLWKqL (grades 4-5), https://bit.ly/3dSusXw (grades 6-7). The endline tests can be accessed at: https://bit.ly/2UtCSx6 (grades 4-5), https://bit.ly/30v0jcU (grades 6-7).

As discussed in the intervention section, control students had access to practice exercises prior to the start of our study. Therefore, the effect that we are estimating is that of shutting down this feature for this set of students for six months. We believe this point estimate is meaningful for two main reasons. First, we are estimating effects on test-score changes, so we can understand how the presence of this feature affects the progress that students make on their math achievement. Second, the practice exercises that control students had previously completed were on different topics and skills, so they should not affect their performance on the topics and skills covered during our study. Nevertheless, to the extent that those practice exercises taught control students skills that carry over to the intervention period, we would be underestimating the effect of the intervention.

4 Results

4.1 Implementation fidelity

The intervention was implemented largely as intended. First, virtually all students across both experimental groups (3,999 out of 4,001 students or 99.9%) logged in at least once to the CAL platform during the evaluation. The average student interacted with the software for 952 minutes during the six months of the study (i.e., about 38 minutes per week, see Figure 2), but usage differed across sites, from 600 minutes (in Ghaziabad, Uttar Pradesh) to 2,417 minutes (in Rajkot, Gujarat). Usage took place mostly at home (instead of during school hours, see Figure A2).

Second, all treatment students completed at least one practice exercise during the study. The average treatment student spent 76 minutes completing practice exercises during the study period (i.e., about 3 minutes per week, see Figure 3). Yet, interaction with the practice exercises also differed across sites, from 43 minutes (in Ghaziabad, Uttar Pradesh) to 130 minutes (in Rajkot, Gujarat).

The practice exercises that treatment students completed covered 18 topics (e.g., "geometry"), 59 subtopics (e.g., "triangles and triangle properties"), and 151 units within those topics (e.g., "classifying triangles based on sides and angles"). The three topics in which students completed the most amount of practice exercises were measurement (14% of all exercises), fractions (10%), and number theory (9.1%).

4.2 Average effects on math achievement

The offer of the intervention had a precisely estimated null effect (of 0.014 standard deviations or SDs) on the math achievement of the average student, regardless of whether we account for students' baseline performance or not (Table 2). In fact, based on the 95% confidence interval, we can rule out effects below -0.035 SDs and above 0.062 SDs. This null average effect is consistent across content domains, with point estimates ranging from 0.002 to 0.008, and across cognitive domains, with point estimates ranging from 0.002 to 0.011 (Table 3).[9] It is also consistent across common items across baseline and endline (which we call "repeated" items) and items that we introduced in the endline (which we call "non-repeated" items; Table A5).

4.3 Heterogeneous effects on math achievement

We explored whether the effect of the intervention differed across the only three student characteristics recorded in our data—sex, enrolled grade, and initial achievement—as we had specified in our pre-analysis plan. We found that the intervention had a moderate-to-large positive effect of 0.14 SDs for students with initially low achievement in math. First, we do so graphically. We plot the effects of the intervention for each within-grade quartile of baseline math achievement (Figure 4). We also present plots of quantile treatment effects by students' percentile on the baseline and endline scores. We examine this heterogeneity in two ways: first, by interacting the treatment indicator variable with each student's (continuous) baseline score and then, by interacting that indicator with indicators for each student's within-grade quartile of baseline achievement (Table 4).

We did not, however, find any evidence of heterogeneous effects by students' sex: female students performed slightly below male students, but the difference was not statistically significant, nor was the interaction between the treatment and female indicator variables (Table A6). We did not find any evidence of heterogeneity in treatment effects by the grade in which students were enrolled. We did not find evidence of heterogeneity by the school that students attended either (Figure A3).

4.4 Average effects on interaction with CAL software

Given that treatment students are required to complete practice exercises at the end of each unit, it is possible that the intervention leads them to potentially completing fewer units in the CAL platform than control students. This would be problematic because, while we expect that practice exercises would *positively* impact the math achievement of treatment students, we would also expect that completing fewer units would *negatively* impact their achievement, and the average effect that we estimate may confound these conflicting influences.

We address this possibility in three ways in Table A7. First, we estimate the effect of the intervention on the number of sessions completed in the CAL platform. Treatment students spent only about 1% more of sessions than control students, but the difference between the two is statistically insignificant. Second, we estimate the effect of the intervention on the total time spent on the platform. Treatment students spent 2.3% more minutes than control students, but again, the

[9] In the reasoning cognitive domain, we find a marginally statistically significant effect, which is likely due to the number of hypothesis tests that we are running (see Table 3).

difference is not statistically significant. Third, we estimate the effect of the intervention on the total time spent on the platform, holding number of sessions completed constant (to obtain an estimate the effect of the intervention on time spent per session). Per session, treatment students spent 1.5% more minutes on the platform, but the difference is small and only marginally statistically significant. In short, we do not see any compelling evidence that the practice exercises held treatment students back.

5 Conclusion

This paper presents one of the first studies that is sufficiently powered to simultaneously rule out meaningful effects of independent practice for the average and detect non-trivial positive effects for lower-performing students. After only six months of an average of three minutes of practice per day, we find that the lowest-performing students attending private schools that cater to relatively well-off families outperformed their control peers by .14 SDs. However, this extra practice had no effect on their average-performing counterparts. These results are robust across sub-group analyses and different schools and cities.

Our study makes an important contribution to three different but related literatures. First and foremost, it identifies an approach to address heterogeneity in students' preparation for schooling, a frontier challenge in developing countries (see Ganimian and Murnane 2016; Glewwe and Muralidharan 2016). This approach demands far less time from students and teachers than remedial interventions with similar effects (e.g., Banerjee et al. 2007b). It can be pursued (mostly) after school hours, without requiring that teachers divert from the curriculum, or that they take time away from core subjects to take students to the computer lab—two factors that have frustrated efforts to scale-up similarly effective interventions (see, e.g., Banerjee et al. 2017; Muralidharan and Singh 2019). And, conditional on the availability of the requisite hardware and software required for the CAL platform in which it is embedded (in this case, Mindspark) it does not require any additional setup costs or training. Given the wide reach of the Mindspark software in India and abroad, and the current need to educate students while they are out of school due to the coronavirus, our results suggest that we could be helping a lot more students catch up with their peers by encouraging independent practice.

Second, our study also contributes to the rapidly evolving body of research on private schools in India and in developing countries more generally. As we mention in the introduction, this literature has consistently found that private schools do not add more value than their public counterparts in math or reading, but that they provide the same results at lower costs (e.g., Muralidharan and Sundararaman 2015; Singh 2015). Our study identifies an approach through which at least unaided private schools could improve learning among low performers. Given the low costs of the underlying platform (see Muralidharan et al. 2019), we are optimistic that a similar approach could be adopted in aided private schools, which make up the bulk of the sector in South Asia (see Andrabi et al. 2007).

Finally, our study offers an important demonstration of how to leverage the growing penetration of educational software products to run rapid cycle randomized evaluations that shed light into the merits of intuitively appealing yet largely untested educational strategies. Perhaps more importantly, it does so in a way that allows researchers to closely monitor students' interaction with the intervention being tested (in this case, independent practice) and to estimate its effect precisely, not only for the average student, but also for relevant sub-groups. We see this as a crucial contribution to research on education technology, given that many interventions that have been evaluated in this space have yielded disappointing results and would benefit from feedback on their effectiveness (Ganimian et al. 2020).

6 References

Andrabi, T., Das, J., Khwaja, A.I., Vishwanath, T., Zajonc, T., 2007. Learning and Educational Achievements in Punjab Schools (LEAPS): Insights to inform the education policy debate. World Bank, Washington, DC.

ASER, 2019. Annual Status of Education Report 2018 (Rural). Provisional Report. Pratham. New Delhi. URL: http://img.asercentre.org/docs/Publications/ASER%20Reports/ASER%202016/aser_2016.pdf.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., Walton, M., 2017. From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application. Journal of Economic Perspectives 31, 73–102. URL: https://www.aeaweb.org/articles?id=10.1257/jep.31.4.73, doi:10.1257/jep.31.4.73.

Banerjee, A.V., Cole, S., Duflo, E., Linden, L., 2007a. Remedying Education: Evidence from Two Randomized Experiments in India. The Quarterly Journal of Economics 122, 1235–1264.

Banerjee, A.V., Cole, S., Duflo, E., Linden, L., 2007b. Remedying education: Evidence from two randomized experiments in India. The Quarterly Journal of Economics 122, 1235–1264.

de Barros, A., Ganimian, A.J., 2020. Are You Smarter than a Fourth Grader? New Evidence on the Math Skills of Indian Children. URL: https://www.dropbox.com/s/u1bcj2zegg3n6jg/de%20Barros%20%26%20Ganimian%202020%20-%20Manuscript.pdf?dl=1.

Bettinger, E., Fairlie, R.W., Kapuza, A., Kardanova, E., Loyalka, P., Zakharov, A., 2020. Does EdTech Substitute for Traditional Learning? Experimental Estimates of the Educational Production Function. Working Paper 26967. National Bureau of Economic Research. Cambridge, MA. URL: http://www.nber.org/papers/w26967, doi:10.3386/w26967. series: Working Paper Series.

Cardelle-Elawar, M., 1990. Effects of Feedback Tailored to Bilingual Students' Mathematics Needs on Verbal Problem Solving. The Elementary School Journal 91, 165–175. URL: https://www.journals.uchicago.edu/doi/abs/10.1086/461644, doi:10.1086/461644. publisher: The University of Chicago Press.

Cardelle-Elawar, M., 1995. Effects of metacognitive instruction on low achievers in mathematics problems. Teaching and Teacher Education 11, 81–95. URL: http://www.sciencedirect.com/science/article/pii/0742051X94000193, doi:10.1016/0742-051X(94)00019-3.

Ganimian, A.J., Hess, F.M., Vegas, E., 2020. Realizing the promise: How can education technology improve learning for all? Technical Report. Brookings Institution. Washington, D.C. URL: https://www.brookings.edu/wp-content/uploads/2020/08/edtech_playbook_full_v2.pdf.

Ganimian, A.J., Murnane, R.J., 2016. Improving Education in Developing Countries: Lessons From Rigorous Impact Evaluations. Review of Educational Research 86, 719–755. URL: https://doi.org/10.3102/0034654315627499, doi:10.3102/0034654315627499.

Glewwe, P., Muralidharan, K., 2016. Improving School Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications, in: Hanushek, E., Machin, S., Woessmann, L. (Eds.), Handbook of the Economics of Education. Elsevier, pp. 653–744.

von Hippel, P.T., Bellows, L., 2018. How much does teacher quality vary across teacher preparation programs? Reanalyses from six states. Economics of Education Review 64, 298–312. URL: https://linkinghub.elsevier.com/retrieve/pii/S0272775717303771, doi:10.1016/j.econedurev.2018.01.005.

Hohn, R.L., Frey, B., 2002. Heuristic Training and Performance in Elementary Mathematical Problem Solving. The Journal of Educational Research 95, 374–380. URL: https://doi.org/10.1080/00220670209596612, doi:10.1080/00220670209596612. publisher: Routledge _eprint: https://doi.org/10.1080/00220670209596612.

Jitendra, A.K., Griffin, C.C., McGoey, K., Gardill, M.C., Bhat, P., Riley, T., 1998. Effects of mathematical word problem solving by students at risk or with mild disabilities. The Journal of Educational Research 91, 345–355.

Jitendra, A.K., Star, J.R., Rodriguez, M., Lindell, M., Someki, F., 2011. Improving students' proportional thinking using schema-based instruction. Learning and Instruction 21, 731–745. URL: http://www.sciencedirect.com/science/article/pii/S0959475211000375, doi:10.1016/j.learninstruc.2011.04.002.

King, A., 1991. Effects of training in strategic questioning on children's problem-solving performance. Journal of Educational Psychology 83, 307–317. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-0663.83.3.307, doi:10.1037/0022-0663.83.3.307.

Kingdon, G.G., 2020. The Private Schooling Phenomenon in India: A Review. The Journal of Development Studies , 1–23URL: https://www.tandfonline.com/doi/full/10.1080/00220388.2020.1715943, doi:10.1080/00220388.2020.1715943.

Kolen, M.J., Brennan, R.L., 2004. Test Equating, Scaling, and Linking. 3rd ed., Springer, New York, NY.

Kramarski, B., Mevarech, Z.R., 2003. Enhancing Mathematical Reasoning in the Classroom: The Effects of Cooperative Learning and Metacognitive Training. American Educational Research Journal 40, 281–310. URL: https://doi.org/10.3102/00028312040001281, doi:10.3102/00028312040001281. publisher: American Educational Research Association.

Lai, F., Luo, R., Zhang, L., Huang, X., Rozelle, S., 2015. Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in migrant schools in Beijing. Economics of Education Review 47, 34–48. URL: https://linkinghub.elsevier.com/retrieve/pii/S027277571500045X, doi:10.1016/j.econedurev.2015.03.005.

Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., Boswell, M., Rozelle, S., 2013. Computer assisted learning as extracurricular tutor? Evidence from a randomised experiment in rural boarding schools in Shaanxi. Journal of Development Effectiveness 52, 208–231.

Lai, F., Zhang, L., Qu, Q., Hu, X., Shi, Y., Boswell, M., Rozelle, S., 2016. Does computer-assisted learning improve learning outcomes? Evidence from a randomized experiment in public schools in rural minority areas in Qinghai, China. Working Paper 237. Stanford University, Freeman Spogli Institute for International Studies. Stanford, CA. URL: https://reap.fsi.stanford.edu/publications/does_computerassisted_learning_improve_learning_outcomes_evidence_from_a_randomized_experiment_in_public_schools_in_rural_minority_areas_in_qinghai_china.

Mevarech, Z.R., Kramarski, B., 2003. The effects of metacognitive training versus worked-out examples on students' mathematical reasoning. British Journal of Educational Psychology 73, 449–471. URL: https://onlinelibrary.wiley.com/doi/abs/10.1348/000709903322591181, doi:10.1348/000709903322591181. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1348/000709903322591181.

Ministry of Law and Justice, 2009. Right of Children to Free and Compulsory Education Act, 2009. URL: https://mhrd.gov.in/sites/upload_files/mhrd/files/upload_document/rte.pdf.

Mo, D., Zhang, L., Wang, J., Huang, W., Shi, Y., Boswell, M., Rozelle, S., 2015. Persistence of learning gains from computer assisted learning: Experimental evidence from China. Journal of Computer Assisted Learning 31, 562–581. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jcal.12106, doi:10.1111/jcal.12106. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcal.12106.

Mullis, I.V.S., Martin, M.O., 2017. TIMSS 2019 Assessment Frameworks. International Association for the Evaluation of Educational Achievement (IEA), Chestnut Hill, MA. URL: http://timss2019.org/wp-content/uploads/frameworks/T19-Assessment-Frameworks.pdf.

Muralidharan, K., Kremer, M., 2008. Public-Private Schools in Rural India, in: Chakrabarti, R., Peterson, P.E. (Eds.), School Choice International: Exploring Public-Private Partnerships. MIT Press, Cambridge, MA, pp. 91–110. URL: https://mitpress.universitypressscholarship.com/view/10.7551/mitpress/9780262033763.001.0001/upso-9780262033763-chapter-5. publication Title: School Choice International Section: School Choice International.

Muralidharan, K., Singh, A., 2019. Improving Schooling Productivity through Computer-Aided Personalization: Experimental Evidence from Rajasthan, RISE Annual Conference 2019, Washington, D.C. URL: https://www.riseprogramme.org/events/rise-annual-conference-2019.

Muralidharan, K., Singh, A., Ganimian, A.J., 2019. Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India. American Economic Review 109, 1426–1460. URL: http://www.aeaweb.org/articles?id=10.1257/aer.20171112, doi:10.1257/aer.20171112.

Muralidharan, K., Sundararaman, V., 2015. The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India. The Quarterly Journal of Economics 130, 1011–1066. URL: http://qje.oxfordjournals.org/lookup/doi/10.1093/qje/qjv013, doi:10.1093/qje/qjv013.

NCERT, 2018a. NAS 2017: Delhi. Report. National Council for Educational Research and Training (NCERT). New Delhi, Delhi. URL: http://www.ncert.nic.in/programmes/NAS/pdf/state/Delhi.pdf.

NCERT, 2018b. NAS 2017: Gujarat. Report. National Council for Educational Research and Training (NCERT). New Delhi, Delhi. URL: http://www.ncert.nic.in/programmes/NAS/pdf/state/Gujarat.pdf.

NCERT, 2018c. NAS 2017: Haryana. Report. National Council for Educational Research and Training (NCERT). New Delhi, Delhi. URL: http://www.ncert.nic.in/programmes/NAS/pdf/state/Haryana.pdf.

NCERT, 2018d. NAS 2017: Tamil Nadu. Report. National Council for Educational Research and Training (NCERT). New Delhi, Delhi. URL: http://www.ncert.nic.in/programmes/NAS/pdf/state/Tamil%20Nadu.pdf.

NCERT, 2018e. NAS 2017: Uttar Pradesh. Report. National Council for Educational Research and Training (NCERT). New Delhi, Delhi. URL: http://www.ncert.nic.in/programmes/NAS/pdf/state/Uttar%20Pradesh.pdf.

NCERT, 2018f. NAS 2017: West Bengal. Report. National Council for Educational Research and Training (NCERT). New Delhi, Delhi. URL: http://www.ncert.nic.in/programmes/NAS/pdf/state/West%20Bengal.pdf.

NIEPA, 2018a. Elementary Education in India: Where Do We Stand? District Report Cards 2016-17. volume 1. National Institute of Educational Planning and Administration, New Delhi, India. Google-Books-ID: mUMmAQAAMAAJ.

NIEPA, 2018b. Elementary Education in India: Where Do We Stand? District Report Cards 2016-17. volume 2. National Institute of Educational Planning and Administration, New Delhi, India. Google-Books-ID: mUMmAQAAMAAJ.

NIEPA, 2018c. U-DISE Flash Statistics 2016-17. National Institute of Educational Planning and Administration, New Delhi, India. URL: http://udise.in/Downloads/Publications/Documents/Flash_Statistics_on_School_Education-2016-17.pdf. google-Books-ID: mUMmAQAAMAAJ.

Pitchford, N.J., 2015. Development of early mathematical skills with a tablet intervention: a randomized control trial in Malawi. Frontiers in Psychology 6. URL: http://journal.frontiersin.org/article/10.3389/fpsyg.2015.00485/abstract, doi:10.3389/fpsyg.2015.00485.

Singh, A., 2015. Private school effects in urban and rural India: Panel estimates at primary and secondary school ages. Journal of Development Economics 113, 16–32. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304387814001175, doi:10.1016/j.jdeveco.2014.10.004.

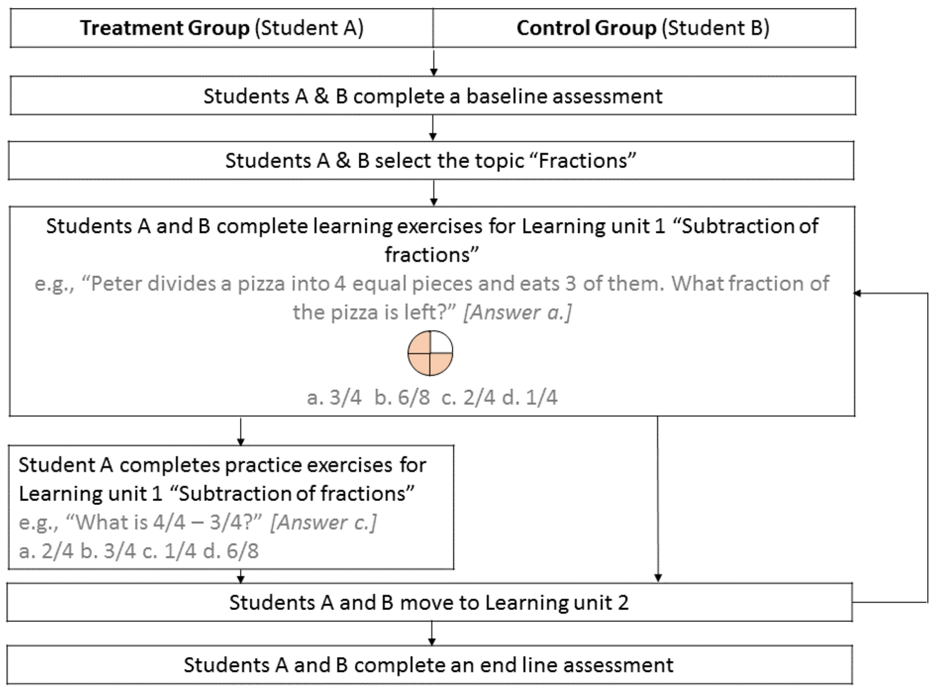
Verschaffel, L., Corte, E.D., Lasure, S., Vaerenbergh, G.V., Bogaerts, H., Ratinckx, E., 1999. Learning to Solve Mathematical Application Problems: A Design Experiment With Fifth Graders. Mathematical Thinking and Learning 1, 195–229. URL: https://doi.org/10.1207/s15327833mtl0103_2, doi:10.1207/s15327833mtl0103_2. publisher: Routledge _eprint: https://doi.org/10.1207/s15327833mtl0103_2.

Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K.R., Ogbuehi, P., 2012. Improving mathematical problem solving in grades 4 through 8: A practice guide. Practice Guide NCEE 2012-4055. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Washington, D.C. URL: http://files.eric.ed.gov/fulltext/ED532215.pdf. publisher: ERIC.

7 Figures and Tables

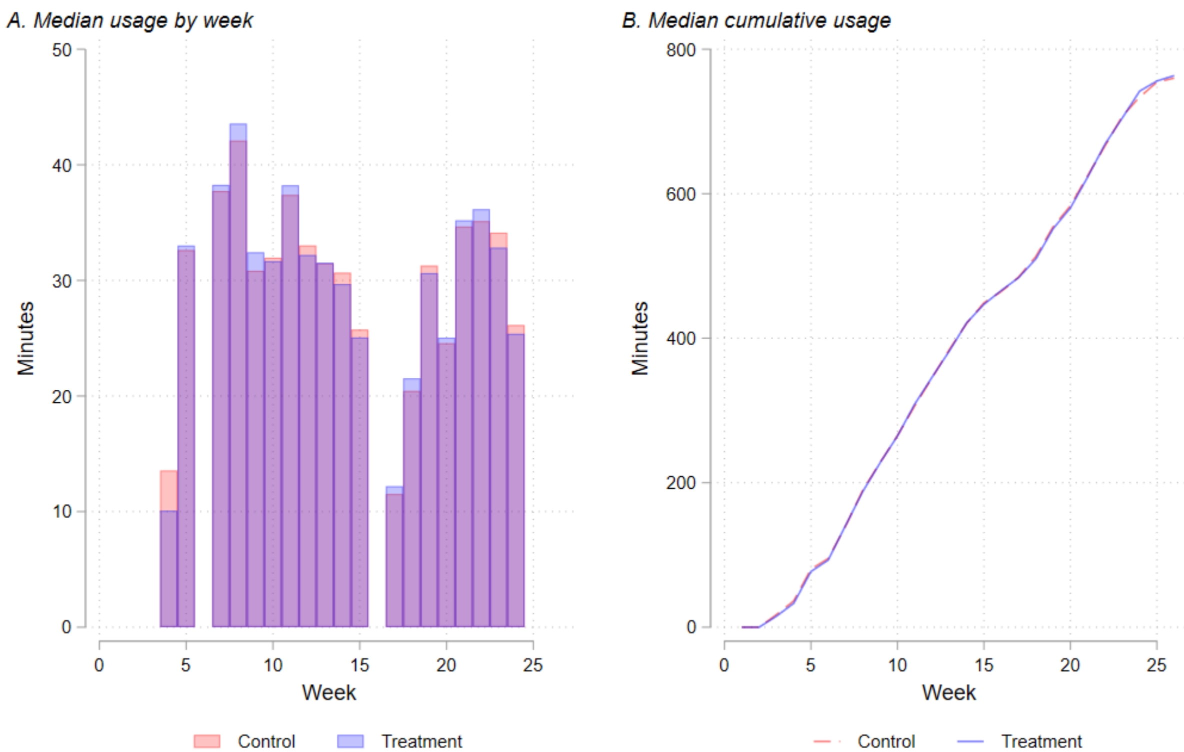
7.1 Figures

Figure 1: *Differences between the sequence of exercises completed by students in the study*



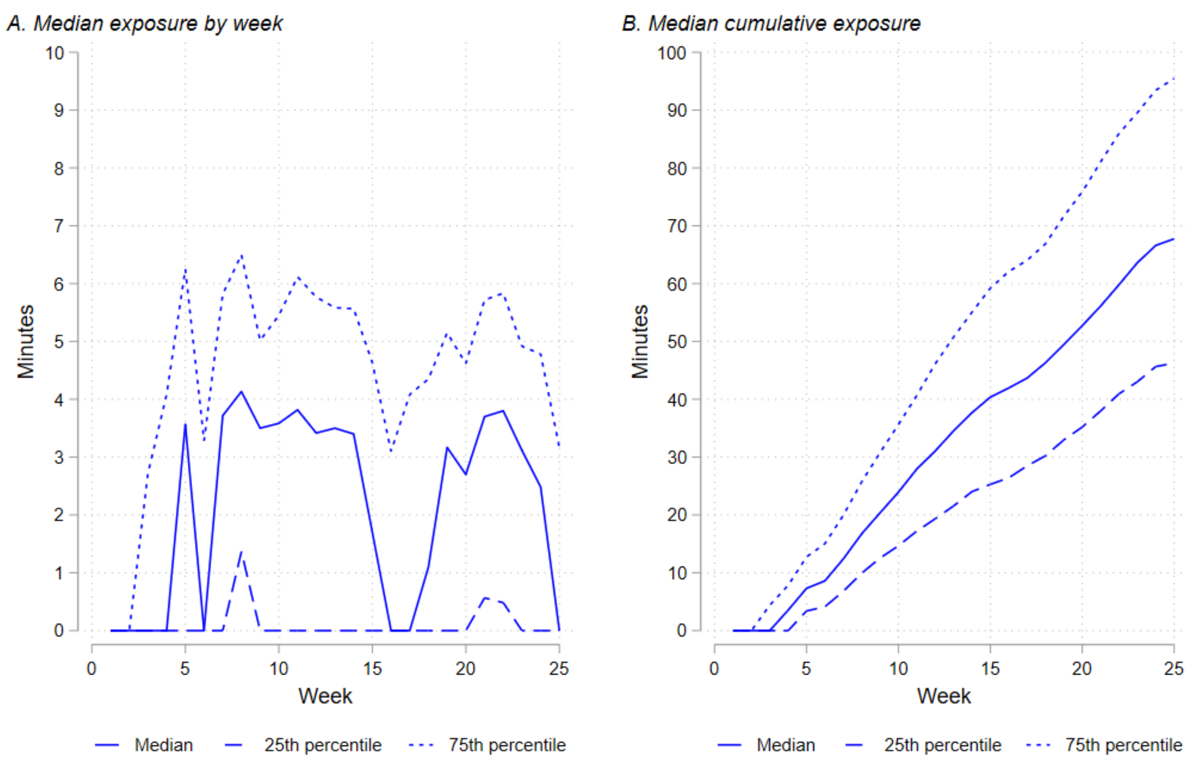
Note: This figure describes the sequence of exercises completed by students in both experimental groups of the study.

Figure 2: *Weekly and cumulative time spent on the CAL platform during the study*



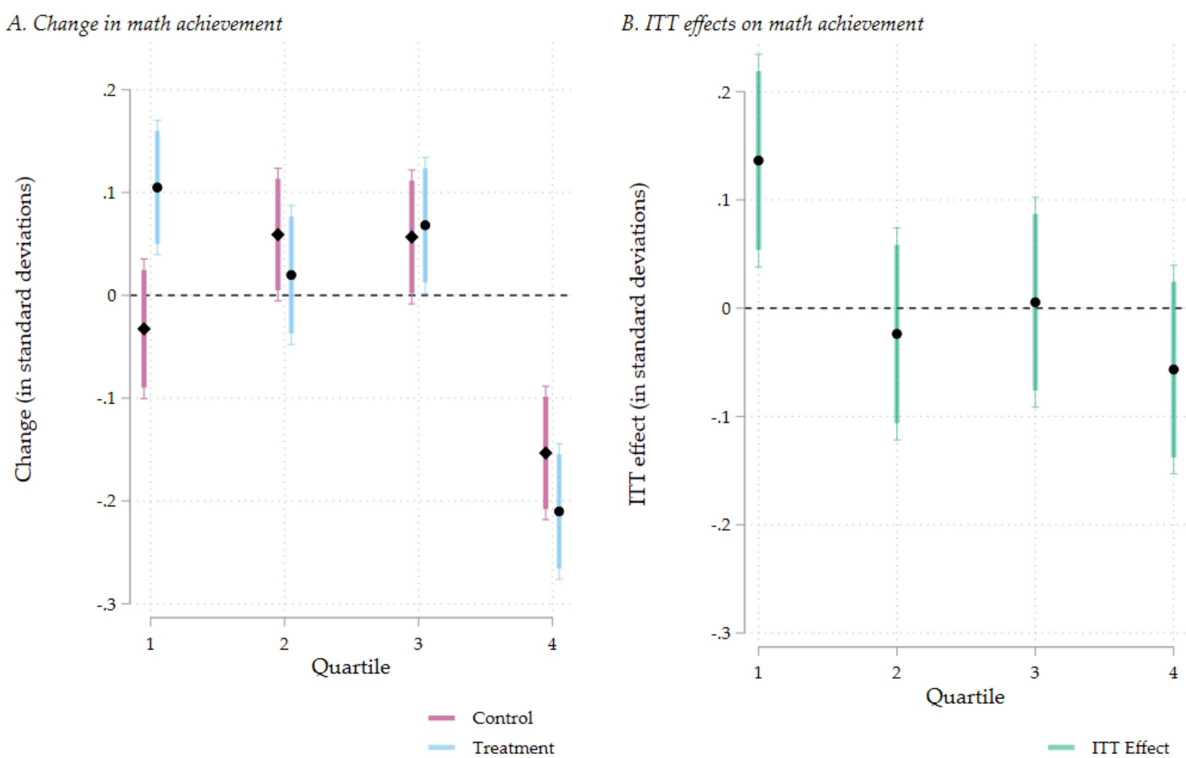
Notes: (1) This figure shows the weekly (panel A) and cumulative (panel B) usage of the CAL platform for the median student, by experimental group. (2) This figure includes all students observed at baseline and endline, regardless of whether they used the software (99.9% of students did). (3) Usage is binned by weeks elapsed since the start of the study (on September 11, 2017).

Figure 3: *Weekly and cumulative time spent on practice exercises during the study among treatment students*



Notes: (1) This figure shows the weekly (panel A) and cumulative (panel B) time spent on practice exercises platform for three groups of treatment students: the median student (i.e., 50th percentile), the 25th percentile, and the 75th percentile, to give a sense of variation in usage in the sample. (2) This figure includes all treatment students observed at baseline and endline, regardless of whether they used the software (99.9% of students did). (3) Usage is binned by weeks elapsed since the start of the study (on September 11, 2017).

Figure 4: *Heterogeneous ITT effects on math achievement at endline, by quartile of baseline performance*



Notes: (1) This figure shows heterogeneity in the change in (panel A) and intent-to-treat (ITT) effect of practice exercises on (panel B) students' achievement in math at endline (after six months), by within-grade quartile of baseline performence. (2) Both panels account for randomization-strata fixed effects. (3) Bars and whiskers show 90-percent and 95-percent confidence intervals, respectively.

7.2 Tables

Table 1: *Balancing checks between experimental groups*

	(1) Control	(2) Treatment	(3) Difference
A. Grade-wise distribution (full sample)			
Grade 4	0.24	0.25	-0.00
	[0.43]	[0.43]	
Grade 5	0.29	0.28	0.01
	[0.45]	[0.45]	
Grade 6	0.25	0.26	-0.01
	[0.43]	[0.44]	
Grade 7	0.22	0.22	0.00
	[0.41]	[0.41]	
B. Balance tests (full sample)			
Baseline score	0.00	-0.03	0.03
	[1.00]	[1.01]	(0.02)
Female	0.53	0.51	0.02
	[0.50]	[0.50]	(0.01)
N (students)	2234	2227	
C. Balance tests (non-attritors)			
Baseline score	0.02	-0.01	0.04
	[1.00]	[1.00]	(0.02)
Female	0.53	0.52	0.01
	[0.50]	[0.50]	(0.01)
N (students)	1984	2017	

Notes: (1) This table compares students in the control and treatment experimental groups on their grade-wise enrollment and characteristics: it shows the mean and corresponding standard deviations for each variable (in brackets) and it compares both groups including randomization-strata fixed effects, showing its mean difference and corresponding standard errors (in parentheses). Panel A compares grade enrollment. It does not perform significance tests because, due to the stratification strategy, grade enrollment is comparable across groups by design. Panel B compares students' baseline score and sex (the only two variables collected at baseline) for all students present at baseline. Panel C does the same only for students who were present at baseline and at endline (90% of the total). (2) ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 2: *ITT effect of practice exercises on math achievement at endline*

	Math (IRT-scaled) score	
	(1)	(2)
Treatment	0.014	0.014
	(0.027)	(0.025)
Baseline score		0.645***
		(0.025)
N (students)	4001	4001
R-squared	0.512	0.588

Notes: (1) This table shows the intent-to-treat (ITT) effect of practice exercises on students' achievement in math at endline (after 6 months). Column 1 shows the simple difference in means; column 2 moreover controls for baseline score. Both estimations include randomization-strata fixed effects. (2) ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 3: *ITT effect of practice exercises on math achievement at endline, by content and cognitive domain*

	A. All students					
	(1) Numbers	(2) Geometry	(3) Data	(4) Knowing	(5) Applying	(6) Reasoning
Treatment	0.002	0.008	0.002	0.004	0.002	0.011*
	(0.004)	(0.005)	(0.006)	(0.005)	(0.004)	(0.006)
Baseline score	0.077***	0.117***	0.114***	0.065***	0.102***	0.132***
	(0.004)	(0.005)	(0.006)	(0.005)	(0.004)	(0.006)
N (students)	4001	4001	4001	4001	4001	4001
R-squared	0.513	0.495	0.405	0.475	0.533	0.501
	B. Low-performing students					
	(1) Numbers	(2) Geometry	(3) Data	(4) Knowing	(5) Applying	(6) Reasoning
Treatment	0.020**	0.051***	0.013	0.027***	0.022**	0.042***
	(0.009)	(0.011)	(0.012)	(0.009)	(0.009)	(0.013)
Baseline score	0.088***	0.125***	0.115***	0.076***	0.110***	0.141***
	(0.007)	(0.009)	(0.010)	(0.008)	(0.007)	(0.011)
N (students)	4001	4001	4001	4001	4001	4001
R-squared	0.517	0.503	0.408	0.478	0.538	0.506

Notes: (1) This table shows the intent-to-treat (ITT) effect of practice exercises on students' (proportion-correct) score in each content (columns 1-3) and cognitive (columns 4-6) domain at endline (after six months). All estimations include randomization-strata fixed effects. Panel A provides average ITT effects among all students. Panel B uses interactions (not shown) to report ITT effects among students in a grade-level's bottom quartile, as per students' performance on the baseline assessment. (2) ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table 4: *Heterogeneous ITT effects on math achievement at endline, by students' baseline performance*

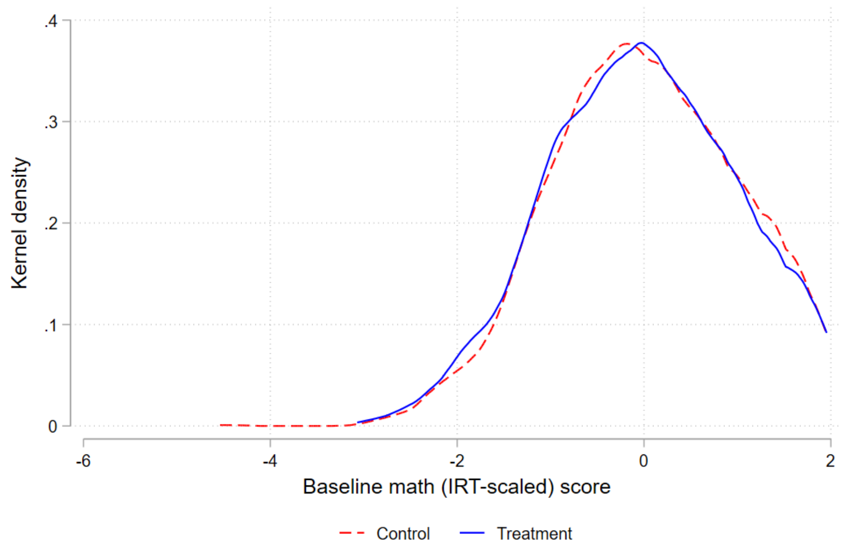
	Math (IRT-scaled) score	
	(1)	(2)
Treatment	0.014	0.136***
	(0.025)	(0.050)
Baseline score	0.672***	0.647***
	(0.028)	(0.042)
Treatment X Baseline	-0.054**	
	(0.025)	
Quartile 2		0.182***
		(0.065)
Quartile 3		0.198**
		(0.085)
Quartile 4		0.144
		(0.110)
Treatment X Quartile 2		-0.160**
		(0.071)
Treatment X Quartile 3		-0.131*
		(0.071)
Treatment X Quartile 4		-0.193***
		(0.070)
N (students)	4001	4001
R-squared	0.588	0.590

Notes: (1) This table shows the intent-to-treat (ITT) effect of practice exercises on students' achievement in math at endline (after six months) by baseline performance, either as a continuous score (column 1) or as a set of quartile indicator variables (column 2). All estimations include randomization-strata fixed effects. (2) ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Appendix A Additional figures and tables

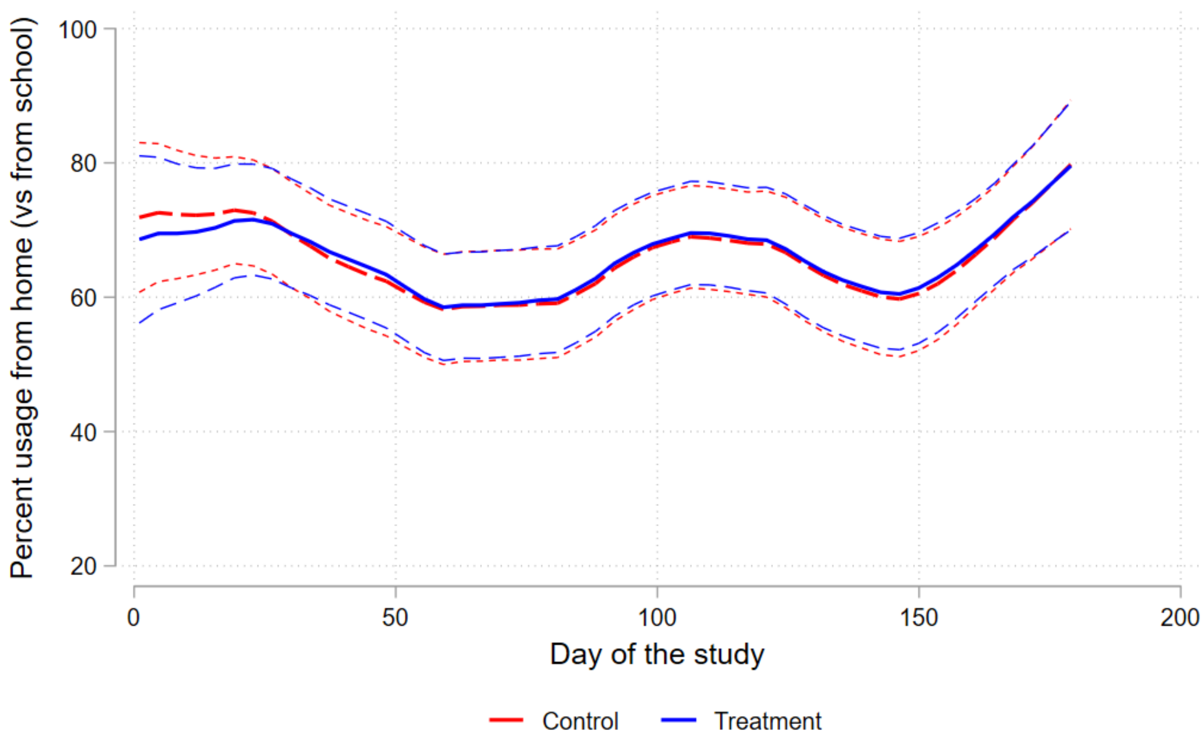
A.1 Additional figures

Figure A1: *Distribution of math (IRT-scaled) scores by experimental group at baseline*



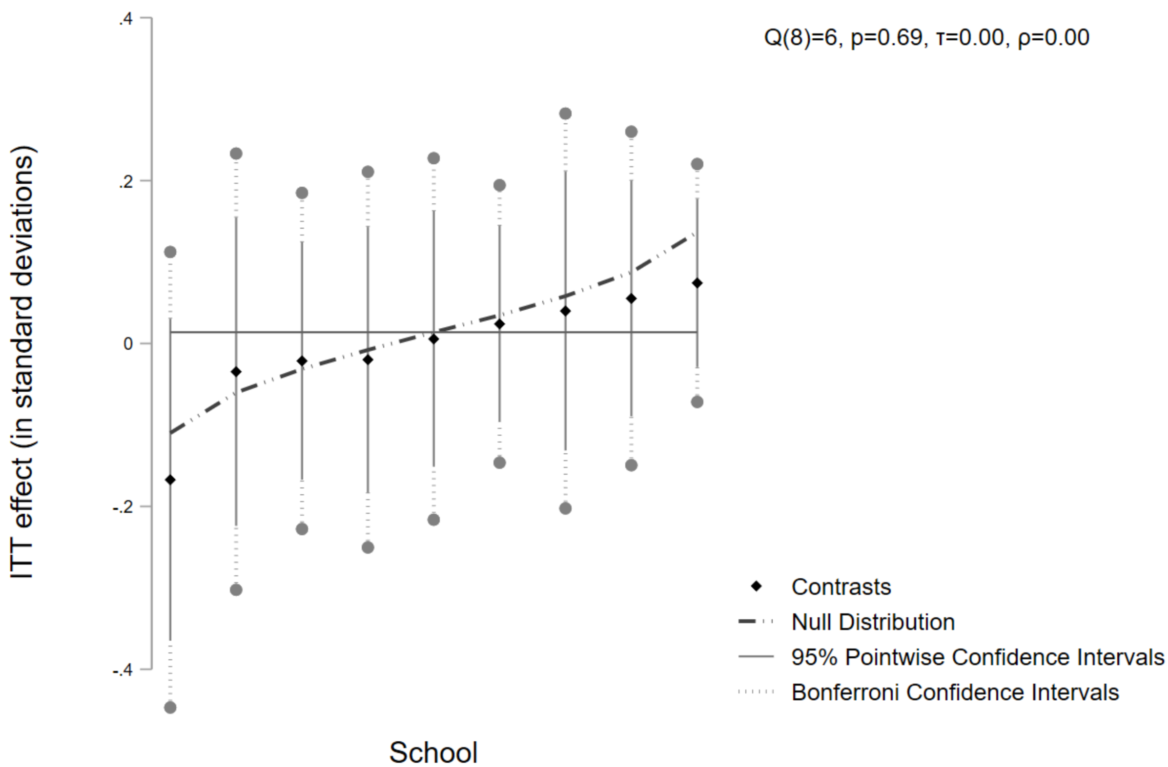
Notes: (1) This figure shows the distribution of scores in the baseline assessment of math for control and treatment students. (2) Scores were scaled using a two-parameter logistic Item Response Theory (IRT) model. (3) This figure includes all students present at baseline and endline.

Figure A2: *Percentage of time spent on the CAL platform at home (instead of at school)*



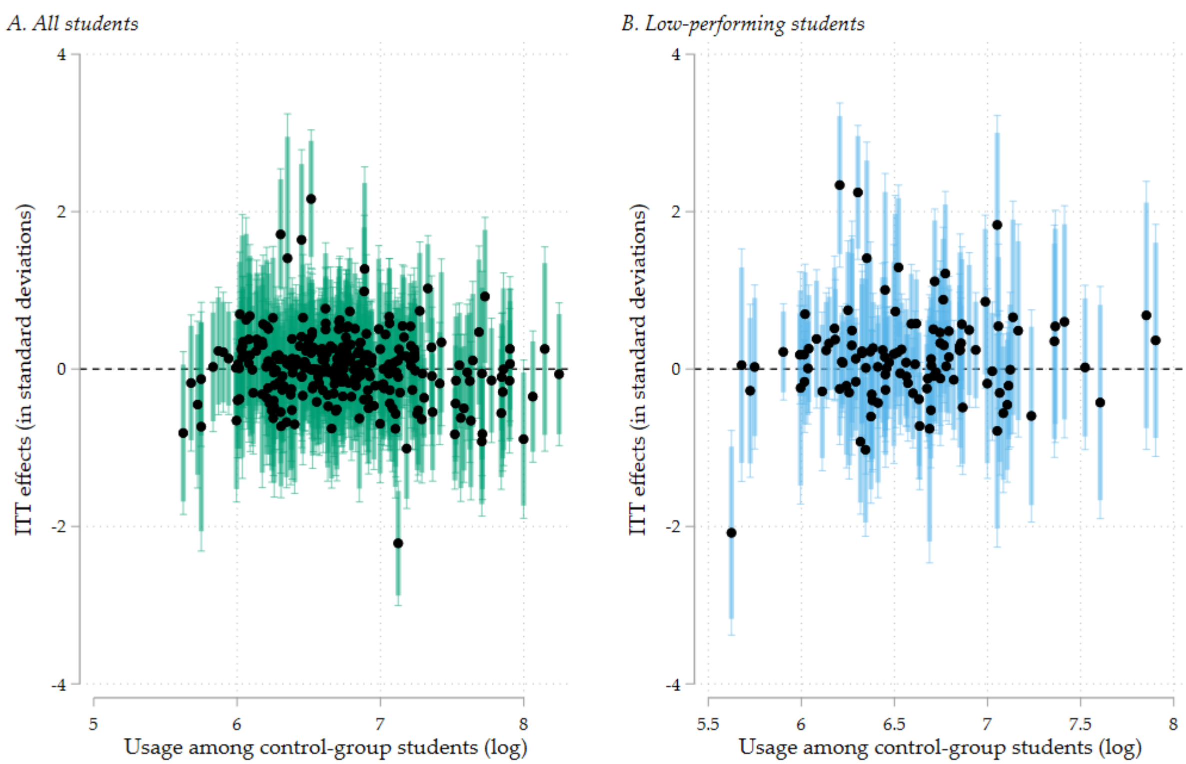
Notes: (1) This figure shows a local polynomial smooth plot with confidence intervals, for the number of minutes that the average student spent using the CAL software at home (rather than during school hours) for each day of the study. (2) 95-percent confidence intervals shown with dashed lines. (3) This figure includes all students who used the software on a given day.

Figure A3: *Heterogeneous ITT effects on math achievement at endline, by school*



Notes: (1) This figure provides a "caterpillar plot" of ITT effects by school (cf. von Hippel and Bellows 2018). Each black dot refers to the point estimate for a given school. (2) Bonferroni confidence intervals adjust standard errors for multiple hypothesis testing. The black solid line shows the null distribution of "effects" that can be expected due to error. τ is the heterogeneity standard deviation. Q refers to Cochran's Q statistic, which follows a χ^2 distribution, and p reports on the corresponding p-value for a test of the null hypothesis of no heterogeneity. ρ estimates the reliability; that is, the share of variance in estimates that is attributable to heterogeneity (rather than error). (3) The estimation controls for student baseline achievement and randomization-strata fixed effects.

Figure A4: *Dose-response relationship*



Notes: (1) This figure shows heterogeneity in the intent-to-treat (ITT) effect of practice exercises on students' achievement in math at endline (after six months) by randomization stratum, for all students (panel A) and students in the bottom quartile of baseline achievement within their grade level (panel B). (2) Bars and whiskers show 90-percent and 95-percent confidence intervals, respectively.

A.2 Additional tables

Table A1: *Overview of primary-school systems across cities in the study, 2018*

	Schools		Students		Efficiency	
	(1) Total	(2) Private	(3) Total	(4) Private	(5) Repetition rate	(6) Dropout rate
Ahmedabad, Gujarat	2,949	1,641 (56%)	986,236	660,776 (67%)	0.65%	-
Faridabad, Haryana	1,236	858 (69%)	337,832	261,779 (77%)	0.51%	-
Ghaziabad, Uttar Pradesh	1,815	1,161 (64%)	484,879	380,561 (78%)	0.08%	4.61%
Kolkata, West Bengal	2,789	645 (23%)	306,156	53,840 (18%)	0.27%	6.90%
New Delhi, Delhi	95	39 (41%)	56,960	31,492 (55%)	0%	-
Rajkot, Gujarat	2,051	1,056 (51%)	426,028	251,263 (59%)	0.40%	0.98%
Tiruchirappalli, Tamil Nadu	2,090	751 (36%)	343,367	184,620 (54%)	0%	-

Sources: NIEPA (2018a,b).

Notes: (1) Primary schools include: primary-only schools; primary schools with upper primary; primary schools with upper primary, secondary, and higher secondary, upper-primary only schools; upper-primary with secondary and higher secondary; primary schools with secondary and higher secondary; and upper-primary schools with secondary. (2) Percentages indicate the share of schools (column 2) and students (column 4) in the private sector. (3) Repetition and dropout rates are not reported separately for private schools. (4) Missing numbers were not reported.

Table A2: *Performance on the national achievement survey of grade 5 across cities in the study, 2017*

	Math		Language		Environmental studies	
	(1)	(2)	(3)	(4)	(5)	(6)
	Government	Government-aided	Government	Government-aided	Government	Government-aided
Ahmedabad, Gujarat	57%	38%	62%	49%	58%	44%
Faridabad, Haryana	52%	-	60%	-	59%	-
Ghaziabad, Uttar Pradesh	54%	48%	61%	61%	60%	54%
Kolkata, West Bengal	53%	61%	64%	70%	59%	65%
New Delhi, Delhi	49%	44%	54%	54%	52%	49%
Rajkot, Gujarat	60%	21%	60%	20%	59%	31%
Tiruchirappalli, Tamil Nadu	39%	38%	52%	54%	45%	43%

Sources: NCERT (2018a,b,c,d,e,f).

Notes: (1) The table indicates the share of test-takers in government (i.e., public) and government-aided (i.e., privately run, publicly subsidized) schools achieving minimum standards in each domain of the reading and math tests of the National Achievement Survey (NAS) of grade 5. (2) No government-aided schools were included in Faridabad.

Table A3: *Share of practice exercises below, at, or above grade level*

	Share of practice exercises...				
Enrolled grade level	...two or more grade levels behind	...one grade level behind	...at grade level	...one grade level above	...two or more grade levels above
Grade 4	0.001	0.070	0.422	0.330	0.177
Grade 5	0.019	0.065	0.407	0.315	0.194
Grade 6	0.010	0.066	0.881	0.042	0.000
Grade 7	0.010	0.281	0.632	0.078	0.000
All grades	0.010	0.113	0.579	0.198	0.099

Notes: (1) This table shows the share of practice exercises completed by treatment students during the study by the grade level in which students were enrolled and the grade level in which each exercise was categorized. Specifically, it shows the share of exercises one or two (or more) grade levels below, at grade level, or one or two (or more) grade levels above the enrolled grade of each student. (2) Practice exercises can be mapped to multiple levels. In this table, if an exercise includes at least the student's enrolled grade level, it is marked as at-level.

Table A4: *Cities, schools, assessment, and software-activation dates*

	(1) School	(2) Baseline date (2017)	(3) Activation date (2017)	(4) Endline date (2018)
Ahmedabad, Gujarat	1	25-26 Sep	6-Oct	3-5 Apr
Faridabad, Haryana	2	26-Sep	6-Oct	27-28 Mar
Ghaziabad, Uttar Pradesh	3	26-Sep	6-Oct	13-14 Apr
Kolkata, West Bengal	4	15-Sep	15-Sep	7-9 Mar
	5	14-Sep	17-Oct	5-7 Mar
New Delhi, Delhi	6	21-22 Sep	17-Oct	16-Apr
	7	25-Sep	6-Oct	9, 11, 12 Apr
Rajkot, Gujarat	8	20-Sep	22-Sep	15-Mar
Tiruchirappalli, Tamil Nadu	9	7-Oct	9-Oct	9, 11-13 Apr

Notes: (1) This table shows the list of sites, schools, assessment, and software activation dates for the study sample. (2) Software activation date refers to the date in which the practice exercises were made unavailable to control students. (3) Schools with multiple baseline and endline dates had multiple grades in the study, which differed on their test dates.

Table A5: *ITT effect of practice exercises on math achievement at endline, by repeated and non-repeated items*

	(1) Repeated items (proportion-correct) score	(2) Non-repeated items (proportion-correct) score
Treatment	0.005	0.002
	(0.004)	(0.005)
Baseline score	0.089***	0.110***
	(0.004)	(0.005)
N (students)	4001	4001
R-squared	0.527	0.491

Notes: (1) This table shows the intent-to-treat (ITT) effect of practice exercises on students' achievement in repeated items across baseline and endline (column 2) and non-repeated items (column 3) domain at endline (after 6 months). Both estimations include randomization-strata fixed effects. (2) ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A6: *Heterogeneous ITT effects on math achievement at endline, by students' sex and enrolled grade*

	Math (IRT-scaled) score	
	(1)	(2)
Treatment	-0.015	0.021
	(0.036)	(0.050)
Female	-0.050	-0.022
	(0.038)	(0.029)
Treatment X Female	0.054	
	(0.050)	
Treatment X Grade 5		-0.011
		(0.068)
Treatment X Grade 6		0.042
		(0.071)
Treatment X Grade 7		-0.067
		(0.073)
Baseline score	0.646***	0.645***
	(0.025)	(0.025)
N (students)	4001	4001
R-squared	0.588	0.588

Notes: (1) This table shows the intent-to-treat (ITT) effect of practice exercises on students' achievement in math at endline (after 25 weeks) for female students (column 1) and students enrolled in different grades (column 2). Both estimations include baseline randomization-strata fixed effects. (2) ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.

Table A7: *ITT effect of practice exercises on usage of CAL platform*

	Number of sessions completed (log)	Total minutes spent on CAL platform (log)	
	(1)	(2)	(3)
Treatment	0.010	0.023	0.015*
	(0.016)	(0.015)	(0.008)
Baseline score	0.068***	0.101***	0.050***
	(0.016)	(0.015)	(0.008)
Number of sessions completed (log)			0.760***
			(0.008)
N (students)	3999	3999	3999
R-squared	0.477	0.504	0.845

Notes: (1) This table shows the intent-to-treat (ITT) effect of practice exercises on the (natural logarithm of) number of sessions that students completed (column 1), on the (natural logarithm of) minutes they spent on the CAL platform (column 2), and on that same number holding the number of sessions completed constant (column 3). All estimations include randomization-strata fixed effects. (2) The estimations excludes 2 (out of 4,001) students who did not spend any time on the software. (3) ***, **, and * indicate significance at the 1, 5, and 10 percent critical level.