NBER WORKING PAPER SERIES

STREAKS TO SUCCESS? THE EFFECTS OF HIGHLIGHTING STREAKS ON STUDENT EFFORT AND LEARNING

Raphaëlle Aulagnon
Julian Cristia
Santiago Cueto
Ofer Malamud

Working Paper 34173
http://www.nber.org/papers/w34173

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2025

This project is the result of a large collaborative effort. We want to especially thank Elena Arias Ortiz, Sheila Avila, Mercedes Mateo, Carolina Méndez, Ben Petrazzini, Alejandra Vargas García and José Zegarra for their support to the project. The project would not have been materialized without the support by a team at the Centro de Investigación Avanzada en Educación of the Universidad de Chile directed by Roberto Araya and the team at the research center GRADE that was coordinated by Natalia Arteta, Jessica Simon and Claudia Sugimaru. Many people at the Ministry of Education of Peru played an essential role for the scale up of the learning platform used in this study including Carmen Beltrán, Sheridan Blossiers, Sandino Canales, Marilú Chávez, Cristina Gálvez, Killa Miranda, Mónica Miyagui and Germán Pachas. We also thank Miguel Chira, Andrea Clavo, Lisseth Escalante, Giacomo Marinelli, Carla de Oliveira and Gera Ríos for outstanding research assistance. We acknowledge excellent comments by David Evans, Stefano Fiorin, Julián Messina and seminar participants at Bocconi University. The ANID/PIA/ Basal Funds for Centers of Excellence (Fund FB0003), the Inter-American Development Bank (including the Korean Fund), the International Research Development Center and the Old Dart Foundation provided funding for the scale-up of the learning platform. The opinions expressed in this document are those of the authors and do not necessarily reflect those of the Inter-American Development Bank. We received approval for this project from the Via Libre (Comite Institutional de Bioetica) ethics board and registered the Conecta Ideas data with the Ministry of Justice in Peru. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Streaks to Success? The Effects of Highlighting Streaks on Student Effort and Learning
Raphaëlle Aulagnon, Julian Cristia, Santiago Cueto, and Ofer Malamud
NBER Working Paper No. 34173
August 2025
JEL No. I21, I25

ABSTRACT

We examine whether highlighting streaks—instances of repeated and consecutive behavior when completing learning tasks—encourages 4th to 6th grade students in Peru to increase their use of an online math platform and improve learning. 60,000 students were randomly assigned to receive messages that (i) highlighted streaks, (ii) provided personalized reminders with positive reinforcement, (iii) provided generic reminders, or (iv) to a control group. Highlighting streaks and providing personalized reminders significantly increased platform use compared to generic reminders and the control group, with streaks more effective on the intensive margin and personalized reminders more effective on the extensive margin. Highlighting streaks also significantly improved math achievement compared to the control group among the 1,500 students who took an endline test, although differences with other treatment arms were not significant.

Raphaëlle Aulagnon
Bocconi University
raphaelle.aulagnon@phd.unibocconi.it

Julian Cristia
Inter-American Development Bank
jcristia@iadb.org

Santiago Cueto
GRADE
and Pontificia Universidad Católica del Perú
scueto@grade.org.pe

Ofer Malamud
Northwestern University
School of Education and Social Policy
and NBER
ofer.malamud@northwestern.edu

1. Introduction

Educational investments often require students to exert effort on tasks that have relatively low returns in the short run, such as paying attention in class or completing extra assignments over the summer (Levitt et al. 2016). Given the barriers that many students encounter when exerting effort in the face of high discount rates, there is growing interest in the use of behavioral strategies for increasing student effort to improve educational outcomes (Koch et al. 2015, Lavecchia et al. 2016). Researchers have explored whether students can be encouraged to exert more effort through financial and non-financial incentives (e.g., Fryer 2011, Bettinger 2012), parental involvement (e.g., Bergman 2019, Berlinski et al. 2021), information about future benefits (e.g., Jensen 2010), personalized reminders (e.g., Castleman and Page 2017), and more. Another approach for encouraging student effort, which has not received much attention in the academic literature, is informing students about their *streaks*—i.e., instances of repeated and consecutive behavior—when completing learning tasks.

Why might informing students about their streaks increase their effort? Several psychological mechanisms have been proposed. Maintaining a streak can become a goal in and of itself, independent of the behavior on which it is based (Silverman and Barasch 2023). This may be amplified by loss aversion, as individuals try to avoid losing their streaks. Streaks also add structure to an activity, which can help simplify thinking and decision-making. This can further serve to "gamify" the activity by creating rules and quantifying the outcome (Weathers and Poehlman 2023). Finally, streaks may provide individuals with a sense of consistency and satisfy an inherent preference for order and uniformity (Silverman and Barasch 2023).

The use of streaks to increase engagement and learning is already common in many online settings, but its impact remains open to debate.[1] For example, streaks are a central feature of Duolingo, arguably the most popular language learning app in the world. An entry in the Duolingo Blog argues that its streak feature is "designed to keep you motivated towards your language learning goals".[2] In contrast, Khan Academy, the online educational platform, opted to retire its streak feature due to worries that it might negatively impact motivation in certain circumstances.[3] However, despite the growing use of streaks and the debates about their possible effects, there is limited empirical evidence examining their causal impacts. Understanding whether gamification features such as streaks improve student engagement and learning will become even more important as technology assumes a greater role in educational service delivery and students' learning experiences.[4]

Our study explores whether highlighting streaks can encourage 4th to 6th grade students in Peru to increase their use of an online math platform and improve their learning over the summer break, a time when students may experience learning loss (Cooper et al. 1996, Atteberry and McEachin 2016). We conducted a randomized field experiment over eight weeks with 60,000 students in households who had downloaded and used the math platform during the preceding academic year. We compare the effect of sending messages which (i) highlighted streaks of completed assignments, to messages which (ii) provided personalized reminders with positive reinforcement, (iii) provided generic reminders, and (iv) to a control group which did not receive

[1] Streaks are also common in other settings: e.g. factories often track their streaks of days without accidents; individuals often track their streaks of avoiding unwanted behaviors such as smoking or consuming alcohol.
[2] See Mansur (2022). Other examples of learning apps that use streaks include Elevate, Memrise, and Codecademy. Streaks are also pervasive on social networking sites like Snapchat and fitness apps like Fitbit.
[3] They explain that "as currently designed, streaks can actually be demotivating, especially when circumstances beyond one's control…can be the reason a streak gets broken." See Khan Academy (2021)
[4] This shift is being accelerated by advances in generative AI for educational applications (Acemoglu et al. 2023).

any messages.[5] The use of personalized and generic reminders as a benchmark for comparison is informative because these types of messages have been employed to address the issue of limited attention and to encourage utilization in various contexts, such as education (e.g., Castleman and Page 2017), savings (e.g. Karlan et al. 2016), and health (e.g., Eleches et al. 2011).

The online math platform in our study, called Conecta Ideas, was aligned with the national math curriculum and provided a weekly set of 30 math exercises. Messages were sent through app notifications over the six weeks of treatment, while a baseline test and an endline test were administered through the app in the weeks before and after the treatment period. Since teachers or schools did not require it, use of the learning platform over the summer was low: only 5.3 percent of students in the control group used the platform at least once, and these students connected for 31.7 percent of weeks in the treatment period.

All our treatment arms generated positive and significant effects on platform use. The largest impacts were observed for highlighting streaks and providing personalized reminders, although they operated on different margins. Personalized reminders generated larger effects than highlighting streaks on the extensive margin of connecting at least once to the platform (3.8 vs. 2.8 percentage points). But highlighting streaks showed larger effects than personalized reminders on the intensive margin, in terms of the fraction of weeks connected conditional on connecting at least once (9.4 vs. 6.9 percentage points). The effects on these two margins counterbalance each other, leading to similar effects of highlighting streaks and providing personalized reminders on the overall fraction of weeks connected and the fraction of total exercises completed. We do not observe a discouragement effect from highlighting streaks.

[5] Note that the Streaks and the Personalized Reminder treatments included two messages per week, whereas Generic Reminders included only one message per week.

We also find suggestive effects on learning. Pooling our treatments, we estimate a 0.10 to 0.12 standard deviation increase in performance on an endline math test that was administered in the week immediately after the six-week treatment period. These pooled effects appear to be driven mostly by the notifications of streaks, the only treatment arm for which the effects on learning are consistently large and significant: students whose streaks were highlighted consistently scored approximately 0.13 to 0.17 standard deviations higher than the control group.

However, the stronger effects of streaks on learning need to be interpreted with caution. First, these effects are not significantly different from those of the other treatment arms (and there are significant effects for the generic reminder treatment in some specifications). Second, as with the low rates of platform use overall, participation in the endline tests was extremely low at only 2.3 percent (approx. 1,500 students). Third, participation in the endline test was significantly higher for students whose streaks were highlighted (by 0.7 percentage points) and students who received positive reinforcement (by 0.4 percentage points) compared to the control group. Finally, while baseline characteristics are balanced between treatment arms and the control group among students taking the endline test, some of the confidence intervals for these differences are sizeable.

To further explore the link between platform use and learning, we also present additional evidence showing that students who connected to the math platform on weeks when specific topics (e.g., probability, geometry) were covered had higher achievement on these topics in the endline test compared to students who connected on weeks that covered other topics.

To our knowledge, this is the first experimental study examining the effects of highlighting streaks in an educational context.[6] The only other study examining the effects of highlighting

[6] There are some proprietary A/B testing of streaks by online platforms, but these are not described in detail. For example, Duolingo reported a 1% increase in daily active users and a 3% increase in Day 14 retention when streaks were highlighted (Econsultancy 2023). They also report that surfacing a "Streak Wager" store item at the end of a lesson led to significant gains in retention, with Day-7 retention increasing by 14% (Duolingo Blog 2023).

streaks on individual behavior is a recent article by Silverman and Barasch (2023). They explore the impact of highlighting intact vs. broken streaks on subsequent engagement in the context of consumer behavior. They conduct a series of experimental "lab" interventions using language learning and word and number games on Mechanical Turk, finding that highlighting "intact" user streaks serves to increase subsequent engagement relative to when "broken" streaks are highlighted. In contrast, our study is a nationwide field experiment intended to improve student learning.

Our study further contributes to the literature on digital learning in and out of school by examining learning in the summer prior to the return of in-person schooling after the Covid-19 pandemic. Prior to the Covid-19 pandemic, Araya et al. (2025) and Muralidharan et al. (2019) showed that computer-aided instruction can lead to improved academic achievement both during school hours and after school. Angrist et al. (2022) delivered math instruction by phone and SMS when schools were closed due to the Covid-19 pandemic and show evidence of positive impacts though other studies found null or even negative effects of similar interventions (Schueler and Rodriguez-Segura 2022, Crawfurd et al. 2023). Less is known about the use of technology to improve educational outcomes over the summer, although previous work has shown that providing books to encourage learning over the summer can be effective (Kim et al. 2006).

2. Context and intervention

2.1. Education in Peru

Primary education in Peru is mandatory and free for all children enrolled in grades one through six. The academic year runs from March to December. Almost all public schools remained closed throughout the academic years of 2020 and 2021 due to the Covid-19 pandemic; students returned to in-person instruction in public schools in 2022. Peru faces significant challenges in math education, which were further exacerbated during the pandemic. From 2019 to 2022, the fraction of Peru's fourth-graders reaching math proficiency fell from 34% to 23%, a decrease attributed to the pandemic's impact (MINEDU 2023).

2.2. The learning platform

The learning platform used in this study, called "Conecta Ideas," included a student app that was developed at the Center for Advanced Research on Education at the Universidad de Chile (and owned by the firm AutoMind). An experimental evaluation in 2017 showed that providing Chilean primary school students with the learning platform twice a week in computer labs increased math achievement by 0.27 standard deviations (Araya et al. 2025). In 2018-2019, the Peruvian research center GRADE adapted the platform for use in schools in Peru. GRADE further modified the program in 2020-2021 to support home-based learning since most public schools in Peru were closed during the pandemic.

The Conecta Ideas app provided students with 30 weekly math exercises that they could complete using a smartphone, tablet, or computer (the vast majority used a smartphone). The platform was free, and students only needed to be connected to the internet once a week to download exercises for the coming week and upload responses from the previous week. Once they logged on, students were presented with 30 math exercises that they needed to solve within the week. After solving each exercise, they received feedback letting them know whether their answer was correct or not.

2.3. The intervention

Our intervention took place during six weeks of the summer break starting from the week of January 17, 2022, and running until the week of February 21, 2022. We also conducted a baseline test and endline test, which took place in the weeks immediately before and after the six-week treatment period. Starting on the Monday of each week during the treatment period, all students logging on to the app were presented with a set of 30 math exercises among four broad topics: (i) numeracy, (ii) geometry, (iii) probability, and (iv) patterns.[7] Students could complete the exercises until the following Sunday. The baseline test and endline test were administered through the app and had a very similar format. They were closely aligned with the exercises included in the six-week treatment period and included 30 exercises covering all the topics presented during the treatment period (i.e., 15 exercises for numeracy, and 5 for geometry, probability, and patterns, respectively).

To encourage the use of the online math platform, notifications were sent through the app to the smartphones of 60,000 households who had used the app during the 2021 academic year. Two app notifications were sent in the week preceding the treatment period (on Monday, January 10, 2022, and Thursday, January 13, 2022) informing parents that students would have the opportunity to do a weekly set of math exercises during the summer, and to encourage students to participate in the baseline test.[8] Two app notifications were also sent to all households in the week following the treatment period (Monday, February 28, 2022, and Thursday, March 3, 2022), encouraging students to complete the endline test. During the six-week treatment period, the

[7] Weeks 2, 3, and 6 covered numeracy; weeks 1, 4, and 5, covered geometry, probability, and patterns respectively.
[8] Appendix A.1 presents the messages sent during the experiment for students in the different treatment arms.

control group did not receive any further notifications, while the three treatment groups received different types of app notifications each week.

The three different types of notifications involved: (i) highlighting streaks of completed assignments, (ii) personalized reminders with positive reinforcement, and (iii) generic reminders. Those in the "generic reminder" treatment received a notification each Monday at 2 pm informing them that a new set of 30 exercises was available to be completed. Those in the "streak" treatment received this Monday notification, plus another notification about their streaks on Thursdays at 2 pm. The exact content of the streak notification varied depending on whether the student had completed the previous week's problem set and the current week's problem set, but always emphasized the benefits of extending or starting a new streak. Those in the "personalized reminder" treatment also received the standard notification every Monday, and another notification on Thursdays at 2 pm. If the student had not completed the current week's problem set, a personalized reminder was sent to encourage them to do so. If the assignment had been completed, a congratulatory message was sent. Thus, this treatment encouraged students who had not yet completed the assignment to complete it and provided positive reinforcement to those who had completed it.

The receipt of notifications may have varied depending on a phone's notification settings. If notifications were enabled for the app, they would appear directly on the phone's screen. If not, the message would be presented when the app was opened. We do not have information on the notification settings for each smartphone used by the students participating in the study, or about whether the student or parents saw the notification.

3. Data and methods

3.1. Data

The primary dataset contains information at the student-exercise level about when each exercise was completed and whether these responses were correct. A complementary file contains the student internal ID and their grade. Student gender was not recorded in the platform, but was assigned for 93% of students using their first name and a prediction algorithm. We also used administrative school-level data reported by the Ministry of Education, containing information on the type of school (multigrade – having teachers responsible for students in more than one grade – or regular schools) and geographic location.

The main outcomes in the study are measures of platform use and student math scores on the endline test. We compute different measures of student platform use using the sample of 60,000 students participating in the study. We also analyze effects on academic achievement, focusing on the subsample of students who completed the endline test. We use data on the endline test to generate a standardized measure of math academic achievement by subtracting the fraction of correct responses from the mean of the control group and then dividing it by the standard deviation.

3.2. Sample selection and randomization

Students participating in the study consisted of a random sample of 60,000 students from the 63,544 students who used the platform at least once during the 2021 academic year. Randomization to treatment was carried out after the baseline test and stratified based on whether the student had completed the baseline test and the number of weeks that the student used the platform in 2021. Forty percent of the sample, or 24,000 students, were assigned to the control group, while 20 percent of the sample, or 12,000 students, were assigned to each of the three treatment groups: Streak, Personalized Reminder, and Generic Reminder. Not surprisingly, given that engagement over the summer break was not required by teachers or schools, participation in the baseline and endline tests was low: 2,268 students took the baseline test and 1,503 took the endline test. We consider the implications of sample selection of taking these tests below.

3.3. Empirical strategy

Our main empirical strategy exploits the randomization across treatment groups to estimate the following equation:

$$Y_{is} = \gamma_0 + \gamma_1 Streak_{is} + \gamma_2 Personalized\ Reminder_{is} + \gamma_3 Simple\ Reminder_{is} + \theta_S + \epsilon_{i_S} \quad (1)$$

where Y_{is} indicates the outcome of interest (a measure of platform use or math achievement) for individual *i* in stratum *s*. The variables *Streak*, *Personalized Reminder*, and *Generic Reminder* are indicators for being assigned to the corresponding treatments and γ_1, γ_2, and γ_3 are the estimated effects of these treatments relative to the control group. θ_S are the strata-level fixed effects and ϵ_{is} is the individual-level error term. In some specifications, we also control for the "baseline" value of the outcome (e.g., average percent of correct responses in 2021).

We also explore how exposure to a particular math topic (i.e. numeracy, geometry, probability, or patterns) affects math achievement in that same topic, exploiting the variation across weeks when different topics were covered since students did not know which topics would be covered each week. Given that students who connected more frequently to the platform would be more likely to be exposed to certain topics, we control for the frequency of platform connections. That is, we compare students who were exposed to certain topics with those who connected on weeks that covered other topics. Specifically, we reshape our data to student-topic level and estimate the following regression:

$$Y_{ij} = \alpha_0 + \alpha_1 Exposed_{ij} + \eta_w + \mu_j + \tau_{ij} \quad (2)$$

where Y_{ij} is standardized achievement in topic j on the endline test for individual i, $Exposed_{ij}$ is a variable representing whether individual *i* was exposed to topic *j* during the treatment period, η_w are fixed effects for the number of weeks that student i used the platform during the treatment period, μ_j are topic fixed-effects and τ_{ij} is the individual-subject error term.[9] The coefficient α_1 represents the effect of being exposed to the math platform on weeks when a tested topic was covered, relative to students who connected on weeks that covered other topics.[10] We cluster standard errors at the student level since there are four observations per student (one for each topic covered).

3.4. Baseline balance

Table 1 examines baseline balance across the four treatment arms for two relevant samples: all students, and students who took the endline test (in panels A and B, respectively). Column 1 presents means for the control group, while columns 2 to 5 show estimated differences for the three individual treatments and the pooled treatment when compared to the control group. Results for the sample of all students show that student characteristics are well balanced between the treatment and control groups. Differences in means between the treatment and control groups are not

[9] For geometry, probability, and patterns, which were covered on specific weeks, the *Exposed* variable is an indicator for whether students connected to the platform during the relevant week. For numeracy, which was covered over three weeks, the *Exposed* variable is the fraction of these weeks that students were connected to the platform.
[10] We also present results for alternative specifications that replace the indicators for number of weeks connected with fixed effects for specific weeks or with student fixed-effects.

statistically significant, except for math achievement in 2021, which shows slightly higher means for the control group. However, the estimated differences, even for this variable, are quantitatively small.

Since only 1,503 students out of 60,000 took the endline test, it is possible that the composition of the sample did not remain balanced across groups once we focus on students that participated in the endline test. However, Panel B of Table 1 suggests that is not the case as there are no statistically significant differences between the control group and any of the treatment groups in the covariates analyzed. In particular, the average fraction of correct answers in 2021 shows good balance between the control and treatment groups.

4. Results

4.1. Effects on platform use

We analyze the effects on platform use for four main outcomes: (i) the fraction of students who connect at least once (the "extensive margin"); (ii) the average weekly connection for those who connect at least once (the "intensive margin"); (iii) the average weekly connection for all students (the overall connection rate);[11] and (iv) the percentage of assigned exercises that are completed. Note that the intensive margin effects are estimated only for those who connect at least once, which varies across treatment arms, so these intensive margin effects should be interpreted with caution.[12]

Table 2 reports our estimated treatment impacts. Since teachers and schools did not require the use of the math learning platform over the summer, take-up was low: only 5.3 percent of

[11] The average weekly connection is the fraction of weeks that students connect during the treatment period.
[12] This is analogous to analyses of labor supply that separately assess changes in the fraction of individuals employed and the number of hours worked conditional on being employed (e.g., Blundell et al., 2013).

students in the control group used the platform at least once, and these students connected 31.7 percent of weeks in the treatment period. The overall fraction of exercises completed was 1.2 percent. In addition to the challenge of inducing primary school students to complete math exercises over the summer, it is also possible that some students did not receive the notifications.[13] Nevertheless, results indicate significant effects of the interventions on take-up and use of the platform.

Focusing on the "extensive margin," Column (1) of Panel A indicates that the Streak treatment increased the likelihood that students connect at least once by 2.8 percentage points relative to the control group, which is lower than the effect of sending personalized reminders (3.8 pp) but higher than the effect of sending generic reminders (1.4 pp). The differences between treatment groups were already present in the first week of the treatment period and continued to grow during the following five weeks (see Panel A of Figure 1 and Panel A of Appendix Table A.1). Panel B of Table 2 shows that pooling the treatments leads to a 2.7 pp higher likelihood of connecting at least once during the treatment period, an increase of 50 percent over the mean of the control group.

Turning to the "intensive margin," Column (2) of Panel A demonstrates that the Streak treatment increased the fraction of weeks connected by 9.4 percentage points relative to the control group, surpassing the effects of sending personalized reminders (6.9 pp) and sending generic reminders (2.5 pp). Panel B of Figure 1 and Appendix Table A.2 show that the effects of the Streak and Personalized Reminder treatments were similar in the first week of the treatment period but start to diverge during the following weeks. This is consistent with the idea that streaks become

[13] Households may have overlooked our notifications, silenced the notifications from our app, or removed the app altogether from their phones since installing it earlier in the year.

relevant once students develop a habit of connecting to the platform and try to avoid breaking longer streaks as they emerge over time. As shown in Panel B of Table 2, pooling the three treatments produces an increase in average weekly connections of 6.5 percentage points, a 21 percent increase over the mean of the control group.

The impacts on the average weekly connection and the fraction of exercises completed, shown in columns (3) and (4), effectively combine the intensive and extensive margins. Here, the Streak and Personalized Reminder treatments generate effects of similar magnitude. While small in absolute levels, these are large effects in relative terms, at over 100 percent of the control group means. This is because the means of the control group are low, between 1 and 2 percent, reflecting the fact that most children did not connect to the platform over the summer.

Results indicate that the Streak treatment generated positive effects on overall platform use, larger than those from Generic Reminders but similar to those from Personalized Reminders. But did the Streak treatment lead to repeated connections in consecutive weeks? To explore this, Appendix Table A.3 presents the effects of the treatments on the fraction of students that used the platform consecutively for up to six weeks. For example, while only 0.5 percent of students in the control group were able to keep up a streak of four weeks, the Streak treatment showed a much higher rate, with 1.5 percent of students in this group achieving a four-week streak. These patterns are also evident when comparing the fraction of students with streaks in the Streak treatment and the Personalized Reminder treatment; we observe a significantly higher likelihood of having five or six consecutive weeks for the Streak treatment compared to the Personalized Reminder treatment.[14]

[14] We observe substantially larger magnitudes when estimating the likelihood of streaks for the sample of students who took the baseline (see Panel A of Appendix Table A.4).

One potential downside of highlighting streaks is the loss in motivation when a streak is broken. To explore this issue, Panel B of Appendix Table A.3 presents the number of consecutive weeks during which a student does not connect. If the Streak treatment discouraged students from connecting again once they disconnected, we might expect to see longer periods when students do not connect at all. Comparing the likelihood of consecutive weeks without a connection in the Streak treatment compared to the Personalized Reminder, we do not find evidence for this possibility.

We also analyze whether the effects are concentrated on the days when messages are sent or whether they generate increases in connections on other days of the week. Appendix Table A.5 documents that effects are concentrated on days when messages are sent (i.e., Monday for the three treatments and Thursdays for Streak and Personalized Reminder). We explore heterogeneous effects on the extensive margin in Appendix Tables A.6 and A.7. While there is variation in treatment impacts across subsamples, we observe that the impacts of Personalized Reminders are consistently larger than both the Streak and the Generic Reminder treatments. Examining heterogeneous effects on the intensive margin in Appendix Tables A.8 and A.9, we also observe that the impacts of the Streaks treatment are consistently larger than both the Personalized Reminders and the Generic Reminder treatments. As we discuss in Appendix A.2 on external validity, the consistent pattern for the extensive and intensive margins across subsamples suggests that the main findings of our study may apply more broadly.

4.2. Effects on taking the endline test

The effects described thus far pertain to platform use during the 6-week treatment period, when treatment students were receiving messages while control students did not. While we cannot assess persistence following the end of treatment due to data limitations, we explore student behavior in the week following treatment when the endline test was administered. During this week, all 60,000 students received two notifications encouraging them to connect to the platform and complete the endline test.

Appendix Table A.10 reveals that students in the Streak treatment had a 0.7 percentage point higher likelihood of taking the endline test compared to students in the control group who had a baseline level of 2.3 percent. The corresponding increase for the Personalized Reminder treatment was 0.4 percentage points, and no effects were found for the Generic Reminder treatment. Among students who took the endline test, those in the Streak treatment also completed a larger number of exercises compared to students in the control group, an effect of 7.7 percentage points on a base of 66.1 percent; no differences were observed for Personalized Reminder and Generic Reminder. Taken together, these results suggest that the Streak treatment generated a higher engagement with the platform in the week immediately after the end of treatment.

4.3. Effects on learning

To examine impacts on learning, we restrict the sample to the 1,503 students who participated in the endline test. As mentioned earlier, students who took the endline test show higher baseline use and math achievement compared to the full sample. Panel B of Table 1 indicates that pre-treatment characteristics are balanced across treatment groups and the control group for this subsample of students. However, the confidence intervals for these differences are quite large.[15] Also, this endline test was closely aligned with the material presented during treatment and does not

[15] For example, we can only reject differences larger than +/– 0.14 of a standard deviation in baseline math scores between the Generic Reminder group and the control group.

represent a broad-based measure of academic achievement. Therefore, these results need to be interpreted with caution.

Table 3 reports the estimated effects of our treatments on endline math achievement using several alternative specifications: based on all 30 exercises (columns 1 and 2) or based on only the exercises attempted (columns 3 and 4); and with (columns 2 and 4) or without (columns 1 and 3) controls for baseline academic achievement in 2021. Across these different specifications, we see that the Streak treatment produced an increase in math academic achievement of 0.13 to 0.17 standard deviation units, while the Personalized Reminder and Generic Reminder treatments generate positive but often insignificant effects on this measure of learning. The differences in impacts between treatments are not statistically significant, and pooling the treatments in Panel B yields an increase of 0.10 to 0.12 standard deviations in the endline test score.

These results are representative of students who completed the endline test, who had higher attachment to the learning platform, and for whom the treatments generated large effects in absolute value compared to the sample of all students participating in the study. In fact, Appendix Table A.11 documents that, for the sample of students that took the endline test, the Streak treatment increased the percent of exercises completed by 21 percentage points and the Pooled treatment by 15 percentage points.

We also derive the effect of platform use on learning by estimating two-stage least squares (2SLS) specifications in which we instrument for the proportion of weeks that students connected to the platform with (i) separate indicators for each treatment arm or (ii) an indicator for the pooled treatments. These results are shown in Appendix Table A.12 and suggest that full exposure to the topics covered on the platform increased achievement by 0.4 to 0.7 standard deviations.[16] To put these results into context, the ITT and TOT effects on math achievement in Muralidharan et al. (2019) are 0.37 and 0.60 standard deviations, respectively, while the ITT effects in Araya et al. (2025) are 0.27 standard deviations.[17] Note that our estimates reflect the effect of the intervention conditional on participation, not the average effect across the full target population. Given low take-up, the true intent-to-treat effects on learning for the overall population would be much smaller.

To further explore the effects of the platform on learning, we leverage variation in the topics to which students were exposed. Table 4 estimates equation (2), showing the effect of being exposed to the math platform on weeks when a specific topic was covered, relative to students who connected on weeks that covered other topics. Students did not know which topics would be covered each week. However, since students who connected more frequently to the platform were more likely to be exposed to certain topics, we control for the intensity of platform connections by adding indicators for the number of weeks students connected (and also add topic fixed effects).

Panel A shows positive and significant effects of exposure to a particular topic during the treatment period on endline performance in that topic (relative to exposure to other topics). Our main estimates, which control for the number of weeks connected (column 1), are robust to replacing the indicators for numbers of weeks connected with fixed effects for: (i) connecting in specific weeks such as week 1 or week 2 (column 2); and (ii) student fixed effects (column 3).

[16] These estimates are essentially scaled up versions of the "reduced form estimates" from Table 3 where the compliance rates are derived from a version of Table 2 restricted to students who participated in the endline test (shown in Appendix Table A.13).
[17] Appendix Table A.14 presents ITT estimates from a broader range of studies examining the effects of learning platforms on academic achievement in developing countries.

In Panel B, we present a placebo exercise in which the dependent variable is math achievement in a topic for the *baseline* test rather than the endline test. Since the baseline test was taken prior to treatment, we should not observe any significant associations between the exposure to a topic during the treatment period and achievement at baseline. Our results indicate no association between exposure to a topic and baseline math achievement in that topic in any of the specifications, indicating that the patterns in Panel A are not due to confounding factors. Appendix A.2 discusses the external validity of the effects on academic achievement.

5. Conclusion

This paper presents the results of a nationwide experimental evaluation in Peru that examined whether highlighting streaks of repeated behavior affected use of an online math platform and subsequent learning. Highlighting streaks generated positive effects on the fraction of students connecting to the platform at least once (the "extensive margin"), albeit lower than those from sending personalized reminders to students. Moreover, highlighting streaks generated larger effects on the fraction of weeks connected among students who connected at least once (the "intensive margin") compared to personalized reminders. Looking at overall measures of platform use that combine both margins, such as the percent of exercises completed, we observe that both highlighting streaks and providing personalized reminders generated similar increases in platform use.

We conclude that highlighting streaks is most effective when the objective is to generate greater attachment among individuals who are already engaged in a desired behavior, though other strategies may be better suited to jump-start a behavior that is not taking place yet. These results also suggest that combining different types of communication over time or across individuals may be ideal. For example, sending personalized reminders may be preferable initially when the goal is to induce individuals to start a desired activity, but highlighting streaks may be a better strategy to encourage higher engagement among those who have already started this activity.

Our study also documents that highlighting streaks produced improvements in math achievement among the 1,500 students who participated in an endline test, though the effects were not statistically different when compared to those from personalized reminders and generic reminders. Additionally, we found that using the platform on weeks when specific topics were covered increased achievement for that topic in the endline test, suggesting that increases in exposure to math do translate to learning gains. These results highlight the potential of behavioral nudges to improve learning among engaged students but also underscore that meaningful gains at scale would require strategies to significantly boost participation.

It is important to acknowledge that our study was implemented over the summer break, and our findings may not generalize to other settings. However, the pattern of results holds for almost all subgroups. Furthermore, while the magnitude of our estimates is somewhat larger for students with higher baseline achievement, these interventions could be targeted at students who are more disadvantaged. More research is needed to document the strength of these results for other populations and other applications, as well as how to best incorporate streaks in optimal communication strategies.

References

Acemoglu, D., D. Autor and S. Johnson. 2023. "Can we have Pro-Worker AI? Choosing a Path of Machines in Service of Minds." Policy Memo, MIT Shaping the Future of Work Initiative. Massachusetts Institute of Technology: Cambridge, MA.

Angrist, N., P. Bergman, and M. Matsheng. 2022. "Experimental Evidence on Learning Using Low-Tech When School Is Out." *Nature Human Behaviour* 6 (7): 941–50. https://doi.org/10.1038/s41562-022-01381-z.

Araya, R., E. Arias Ortiz, N. Bottan, and J. Cristia. 2019. "Does Gamification in Education Work? Experimental Evidence from Chile." IDB Working Paper No. 304. Inter-American Development Bank, Washington, DC. https://doi.org/10.18235/0001777.

Araya, R., E. Arias Ortiz, N. Bottan, and J. Cristia. 2025. "Integrating Learning Platforms within Regular School Time: Experimental Evidence from Chilean Primary Schools." *Economics of Education Review* 106: 102647. https://doi.org/10.1016/j.econedurev.2025.102647.

Bai, Y., D. Mo, L. Zhang, M. Boswell, and S. Rozelle. 2016. "The Impact of Integrating ICT with Teaching: Evidence from a Randomized Controlled Trial in Rural Schools in China." *Computers and Education* 96: 1–14. https://doi.org/10.1016/j.compedu.2016.02.005

Bai, Y., B. Tang, B. Wang, L.Z. Di Mo, S. Rozelle, E. Auden, and B. Mandell. 2018. "Impact of Online Computer-assisted Learning on Education: Evidence from a Randomized Controlled Trial in China." REAP Working Paper. Rural Education Action Program, Stanford University

Bergman, P. 2019. "How Behavioral Science Can Empower Parents to Improve Children's Educational Outcomes." *Behavioral Science & Policy* 5 (1): 53–67. https://doi.org/10.1177/237946151900500105.

Berlinski, S., M. Busso, T. Dinkelman, and C. Martínez A. 2021. "Reducing Parent-School Information Gaps and Improving Education Outcomes: Evidence from High-Frequency Text Messages." NBER Working Paper No. 28581. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w28581.

Bettinger, E. 2012. "Paying to Learn: The Effect of Financial Incentives on Elementary School Test Scores." *Review of Economics and Statistics* 94 (3): 686–98. https://doi.org/10.1162/REST_a_00217.

Bettinger, E., R.W. Fairlie, A. Kapuza, E. Kardanova, P. Loyalka, and A. Zakharov. 2022. "Does EdTech Substitute for Traditional Learning? Experimental Estimates of the Educational Production Function." *Journal of Labor Economics* 40 (3): 737-77. https://doi.org/10.1086/717727.

Büchel, K., M. Jakob, C. Kühnhanss, D. Steffen, and A. Brunetti. 2020. "The Relative Effectiveness of Teachers and Learning Software: Evidence from a Field Experiment in El

Salvador." Discussion Paper No. 20-06. Department of Economics, University of Bern. https://hdl.handle.net/10419/228894.

Cardim, J., T. Molina-Millán and P. Vicente. 2023. "Can Technology Improve the Classroom Experience in Primary Education? An African Experiment on a Worldwide Program." *Journal of Development Economics:* 164: 103145.

Castleman, B., and L. Page. 2017. "Parental Influences on Postsecondary Decision Making: Evidence From a Text Messaging Experiment." *Educational Evaluation and Policy Analysis* 39 (2): 361–77. https://doi.org/10.3102/0162373716687393.

Crawfurd, L., D. Evans, S. Hares and J. Sandefur. 2023. "Live Tutoring Calls Did not Improve Learning During the COVID-19 Pandemic in Sierra Leone." *Journal of Development Economics* 164: 103114. https://doi.org/10.1016/j.jdeveco.2023.103114.

Pop-Eleches C, Thirumurthy H, Habyarimana JP, Zivin JG, Goldstein MP, de Walque D, MacKeen L, Haberer J, Kimaiyo S, Sidle J, Ngare D, Bangsberg DR. Mobile phone technologies improve adherence to antiretroviral treatment in a resource-limited setting: a randomized controlled trial of text message reminders. *AIDS*. 2011 Mar 27;25(6):825-34. https://doi.org/10.1097/QAD.0b013e32834380c1.

Fafchamps, M., and D. Mo. 2018. "Peer Effects in Computer-assisted Learning: Evidence from a Randomized Experiment." *Experimental Economics* 21(2): 355–82. https://doi.org/10.1007/s10683-017-9538-z.

Ferman, B., L. Finamor, and L. Lima. 2019. "Are Public Schools Ready to Integrate Math Classes with Khan Academy?" MPRA Paper, University Library of Munich.

Fryer, R. 2011. "Financial Incentives and Student Achievement: Evidence from Randomized Trials." *The Quarterly Journal of Economics* 126 (4): 1755–98. https://doi.org/10.1093/qje/qjr045.

Ito, H., K. Kasai, H. Nishiuchi and M. Nakamuro. 2021. "Does Computer-aided Instruction Improve Children's Cognitive and Noncognitive Skills? *Asian Development Review* 38: 98-118.

Jensen, R. 2010. "The (Perceived) Returns to Education and the Demand for Schooling." *Quarterly Journal of Economics* 125 (2): 515–48. https://doi.org/10.1162/qjec.2010.125.2.515.

Karlan, D., McConnell, M., Mullainathan, S., & Zinman, J. (2016). Getting to the Top of Mind: How Reminders Increase Saving. *Management Science*, 62(12), 3393–3411.
Khan Academy. 2021. "Update: Streaks are going away on January 4, 2021". Khan Academy Help Center. https://support.khanacademy.org/hc/en-us/community/posts/360075847492-Update-Streaks-are-going-away-on-January-4-2021

Kim, J. 2006. “Effects of a Voluntary Summer Reading Intervention on Reading Achievement: Results From a Randomized Field Trial.” *Educational Evaluation and Policy Analysis* 28 (4): 335–55. https://doi.org/10.3102/01623737028004335.

Koch, A., J. Nafziger, and H. Nielsen. 2015. “Behavioral Economics of Education.” *Journal of Economic Behavior and Organization* 115 (July): 3–17. https://doi.org/10.1016/j.jebo.2014.09.005.

Lai, F., L. Zhang, Q. Qu, X. Hu, Y. Shi, M. Boswell, and S. Rozelle. 2012. “Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Public Schools in Rural Minority Areas in Qinghai, China.” Rural Education Action Project Working Paper No. 237.

Lai, F., L. Zhang, X. Hu, Q. Qu, Y. Shi, Y. Qiao, and S. Rozelle. 2013. “Computer-assisted Learning as Extracurricular Tutor? Evidence from a Randomised Experiment in Rural Boarding Schools in Shaanxi.” *Journal of Development Effectiveness* 5(2): 208–31. https://doi.org/10.1080/19439342.2013.780089.

Lai, F., R. Luo, L. Zhang, X. Huang, and S. Rozelle. 2015. “Does Computer-assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Migrant Schools in Beijing.” *Economics of Education Review* 47: 34–48. https://doi.org/10.1016/j.econedurev.2015.03.005.

Lai, F., L. Zhang, Y. Bai, C. Liu, Y. Shi, F. Chang and S. Rozelle. 2016. “More Is not Always Better: Evidence from a Randomised Experiment of Computer-assisted Learning in Rural Minority Schools in Qinghai.” *Journal of Development Effectiveness* 8: 449-472.

Lavecchia, A., H. Liu, and P. Oreopoulos. 2016. “Behavioral Economics of Education: Progress and Possibilities.” In *Handbook of the Economics of Education*, 5:1–74. Elsevier. https://doi.org/10.1016/B978-0-444-63459-7.00001-4.

Levitt, S., J. List, and S. Sadoff. 2016. “The Effect of Performance-Based Incentives on Educational Achievement: Evidence from a Randomized Experiment.” NBER Working Paper 22107. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w22107.

Ma, Y., Fairlie, R. W., Loyalka, P., and Rozelle, S. 2020. “Isolating the ‘Tech’ from Edtech: experimental evidence on computer assisted learning in China.” NBER Working Paper No. 26953. National Bureau of Economic Research, Cambridge, MA. https://www.nber.org/papers/w26953.

Mansur, O. 2022. “The habit-building research behind your Duolingo streak”. Duolingo Blog. https://blog.duolingo.com/how-duolingo-streak-builds-habit/

MINEDU. 2023. “Evaluación Muestral de Estudiantes 2022. Resultados de La Evaluación Nacional de Logros de Aprendizaje.”

http://umc.minedu.gob.pe/wp-content/uploads/2023/06/Reporte-Nacional-EM-2022.pdf.

Mo, D., L. Zhang, R. Luo, Q. Qu, W. Huang, J. Wang and S. Rozelle. 2014. "Integrating Computer-assisted Learning into a Regular Curriculum: Evidence from a Randomised Experiment in Rural Schools in Student Engagement and Motivation. Shaanxi." *Journal of Development Effectiveness* 6(3): 300–23. https://doi.org/10.1080/19439342.2014.911770.

Mo, D., Y. Bai, Y. Shi, C. Abbey, L. Zhang, S. Rozelle, and P. Loyalka. 2020. "Institutions, Implementation, and Program Effectiveness: Evidence from a Randomized Evaluation of Computer-assisted Learning in Rural China." *Journal of Development Economics* 146: 102487. https://doi.org/10.1016/j.jdeveco.2020.102487.

Muralidharan, K., A. Singh, and A. J. Ganimian. 2019. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." *American Economic Review* 109 (4): 1426–60. https://doi.org/10.1257/aer.20171112.

Levesque, K., Bardack, S., Chigeda, A., Bahlibi, A., & Winiko, S. 2022. "Impacts of a 2-Year Education Technology Program on Early Primary Learning in Malawi amid Disruptions due to COVID-19." Imagine Worldwide Research Brief No. 1. May (updated October 2022). https://www.imagineworldwide.org/wp-content/uploads/2-Year-Malawi-RCT-Research-Brief-1-Final.pdf.

Schueler, B. and D. Rodriguez-Segura. 2023. "A Cautionary Tale of Tutoring Hard-to-Reach Students in Kenya." *Journal of Research on Educational Effectiveness* 16(3): 442-472. https://doi.org/10.1080/19345747.2022.2131661.

Silverman, J., and A. Barasch. 2023. "On or Off Track: How (Broken) Streaks Affect Consumer Decisions." *Journal of Consumer Research* 49 (6): 1095–1117. https://doi.org/10.1093/jcr/ucac029.

Weathers, D., and T.A. Poehlman. 2023. "Defining, and understanding, commitment to, activity streaks." *Journal of the Academy of Marketing Science*. https://doi.org/10.1007/s11747-023-00944-4

Table 1: Baseline Balance

		Differences				
	Control	Streak	Personalized Reminder	Generic Reminder	Pooled	*N*
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: All students						
School characteristics						
Multigrade	5.91	0.02	0.26	0.22	0.17	*60,000*
		(0.26)	(0.27)	(0.27)	(0.20)	
Located in Lima	49.05	-0.30	0.10	0.78	0.19	*60,000*
		(0.55)	(0.55)	(0.56)	(0.41)	
Student characteristics						
Female	51.12	-0.97	-0.82	0.27	-0.51	*55,302*
		(0.58)*	(0.58)	(0.58)	(0.43)	
Fourth grade	34.55	0.15	0.45	0.30	0.30	*60,000*
		(0.53)	(0.53)	(0.53)	(0.40)	
Fifth grade	31.75	0.10	-0.01	-0.68	-0.20	*60,000*
		(0.52)	(0.52)	(0.52)	(0.39)	
Weeks connected in 2021	5.47	-0.01	-0.00	-0.01	-0.00	*60,000*
		(0.01)	(0.01)	(0.01)	(0.00)	
Math achievement in 2021	0.00	0.00	-0.02	-0.02	-0.02	*59,093*
		(0.01)	(0.01)*	(0.01)**	(0.01)**	
Panel B: Students who took the endline test						
School characteristics						
Multigrade	3.64	0.36	0.42	0.42	-0.38	*1,503*
		(1.41)	(1.47)	(1.47)	(1.27)	
Located in Lima	51.09	2.55	-4.20	-4.20	-0.48	*1,503*
		(3.50)	(3.70)	(3.70)	(3.35)	
Student characteristics						
Female	55.43	0.45	-2.30	-2.30	1.47	*1,400*
		(3.69)	(3.83)	(3.83)	(3.55)	
Fourth grade	48.00	2.85	2.08	2.08	0.98	*1,503*
		(3.54)	(3.79)	(3.79)	(3.41)	
Fifth grade	34.18	-2.85	-2.58	-2.58	-2.68	*1,503*
		(3.31)	(3.48)	(3.48)	(3.17)	
Weeks connected in 2021	10.22	0.07	-0.05	-0.03	-0.01	*1,503*
		(0.06)	(0.05)	(0.06)	(0.04)	
Math achievement in 2021	0.36	-0.05	-0.03	0.00	-0.03	*1,503*
		(0.06)	(0.06)	(0.07)	(0.05)	

Notes: This table presents statistics and estimated differences between the control and treatment groups. Panel A reports statistics for all students participating in the study. Panel B presents statistics for the subsample of students who took the endline test. Columns (1) presents means for the control group. Columns (2) to (5) present estimated coefficients and standard errors from OLS regressions that include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, **, and *, respectively.

Table 2: Effects on Platform Use

	Connected at least once (%)	Average weekly connection (for students that connected at least once, %)	Average weekly connection (for all students, %)	Average exercises attempted (%)
	(1)	(2)	(3)	(4)
Panel A: Individual treatments				
Streak	2.82***	9.36***	1.55***	1.24***
	(0.26)	(1.04)	(0.12)	(0.10)
Personalized Reminder (PR)	3.79***	6.91***	1.65***	1.19***
	(0.27)	(0.95)	(0.11)	(0.10)
Generic Reminder (GR)	1.40***	2.51**	0.56***	0.41***
	(0.24)	(0.98)	(0.10)	(0.08)
p-value (Streak = PR)	0.00	0.03	0.50	0.69
p-value (Streak = GR)	0.00	0.00	0.00	0.00
p-value (PR = GR)	0.00	0.00	0.00	0.00
Panel B: Pooled treatment				
Pooled treatment	2.67***	6.50***	1.26***	0.95***
	(0.18)	(0.76)	(0.08)	(0.07)
N	*60,000*	*4,146*	*60,000*	*60,000*
Mean of the control group	5.30	31.69	1.68	1.17

Notes: This table presents estimated effects of the individual and pooled treatments on different measures of platform use. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate the dependent variable in the regression. The dependent variable in column (1) is an indicator that equals one if the student connected at least once during the six-week treatment period. The dependent variable in column (2) and (3) corresponds to the percent of weeks that the student connected during the six-week treatment period. The dependent variable in column (4) corresponds to the percent of exercises attempted out of all the 180 exercises provided to students during the six-week treatment period. The sample in the regressions reported in columns (1), (3) and (4) includes all students. The sample in the regression reported in column (2) includes only students who connected at least once during the six-week treatment period. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table 3: Treatment Effects on Math Achievement

	(1)	(2)	(3)	(4)
Panel A: Individual treatments				
Streak	0.16**	0.17***	0.13**	0.15***
	(0.07)	(0.06)	(0.06)	(0.05)
Personalized Reminder (PR)	0.05	0.09	0.04	0.07
	(0.07)	(0.06)	(0.06)	(0.06)
Generic Reminder (GR)	0.07	0.08	0.13*	0.13**
	(0.07)	(0.06)	(0.07)	(0.06)
p-value (Streak = PR)	0.14	0.23	0.18	0.16
p-value (Streak = GR)	0.22	0.20	0.96	0.76
p-value (PR = GR)	0.85	0.87	0.23	0.34
Panel B: Pooled treatment				
Pooled treatment	0.10*	0.12***	0.10**	0.12***
	(0.05)	(0.04)	(0.05)	(0.04)
N	*1,503*	*1,503*	*1,503*	*1,503*
Mean of the control group	0.00	0.00	0.00	0.00
Questions in test included	All	All	Attempted	Attempted
2021 controls	N	Y	N	Y

Notes: This table presents estimated effects of the individual and pooled treatments on math achievement. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. Math achievement in columns (1) and (2) is computed considering all 30 questions included in the endline test. Math achievement in columns (3) and (4) is computed considering only the questions attempted by the student (i.e. questions not answered are not included to compute the percent of correct answers). The sample in each regression includes the students who took theendline test. All regressions include strata fixed-effects. Regressions reported in columns (2) and (4) also control for average math performance in 2021. Math achievement has been normalized by subtracting the mean and dividing by the standard deviation of the control group. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

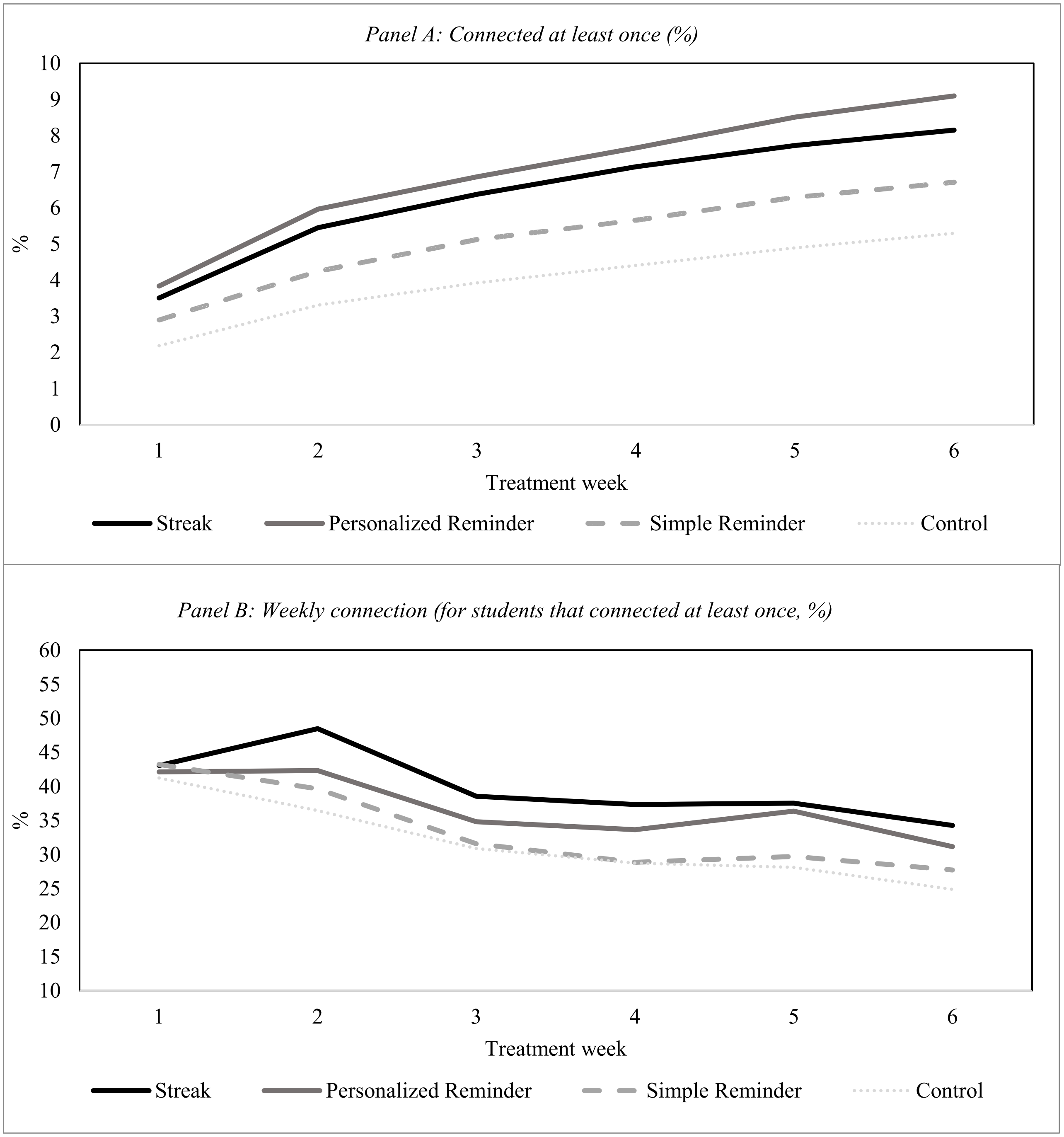
Table 4: Association of Exposure to a Topic and Math Achievement

	(1)	(2)	(3)
Panel A: Association with endline test			
Connected for the tested topic	0.15***	0.16***	0.14***
	(0.04)	(0.04)	(0.05)
N	*3,740*	*3,740*	*3,740*
Topic fixed-effects	Y	Y	Y
Number of weeks connected fixed-effects	Y	N	N
Specific weeks fixed-effects	N	Y	N
Student fixed-effects	N	N	Y

Panel B: Association with baseline test			
Connected for the tested topic	-0.01	-0.00	-0.03
	(0.04)	(0.04)	(0.04)
N	*4,624*	*4,624*	*4,624*
Topic fixed-effects	Y	Y	Y
Number of weeks connected fixed-effects	Y	N	N
Specific weeks fixed-effects	N	Y	N
Student fixed-effects	N	N	Y

Notes: This table presents regression results of the association between exposure to a specific topic and math achievement in that topic in the endline and baseline tests. During the six-week treatment, each week focused on a specific topic: geometry in week 1, numeracy in weeks 2, 3 and 6, probability in week 4, and patterns in week 5. Each column in a panel corresponds to a separate OLS regression. The unit of observation in all regressions is student-subject. That is, there are four observations for each student. In the first observation, the dependent variable is math achievement in geometry and the key independent variable is a dummy that equals one if the student connected to the platform the week when this topic was covered (week 1). Similar observations are constructed for numeracy, patterns and probability. However, since numeracy was covered in multiple weeks, the exposure variable in this case is the share of weeks that the student connected when this topic was covered. The dependent variable in Panel A is math achievement in the endline test. The dependent variable in Panel B is math achievement in the baseline test. All regressions reported in columns (1) to (3) include topic fixed-effects. Regressions reported in column (1) to (3) include also respectively number of weeks connected, student fixed-effects, and specific week fixed-effects. Math achievement for each subject has been normalized by subtracting the mean and dividing by the standard deviation of the control group. Standard errors, reported in parentheses, are clustered at the student level. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Figure 1: Effects on Platform on Use per Treatment Week



Notes: Panel A presents the fraction of students that connected at least once since the start of the experiment by treatment week for each treatment arm. Panel B displays the fraction of students that connected to the platform in a particular week for the different treatment arms including only students that connected at least once during the six-week treatment period.

Table A.1: Effects on Connected At Least Once by Week

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Individual treatments						
Streak	1.30***	2.12***	2.43***	2.70***	2.80***	2.82***
	(0.17)	(0.21)	(0.23)	(0.24)	(0.25)	(0.26)
Personalized Reminder (PR)	1.64***	2.64***	2.92***	3.23***	3.60***	3.79***
	(0.18)	(0.22)	(0.23)	(0.25)	(0.26)	(0.27)
Generic Reminder (GR)	0.71***	0.93***	1.19***	1.24***	1.39***	1.40***
	(0.16)	(0.20)	(0.21)	(0.22)	(0.24)	(0.24)
p-value (Streak = PR)	0.12	0.05	0.08	0.07	0.01	0.00
p-value (Streak = GR)	0.00	0.00	0.00	0.00	0.00	0.00
p-value (PR = GR)	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: Pooled treatment						
Pooled treatment	1.22***	1.89***	2.18***	2.39***	2.60***	2.67***
	(0.12)	(0.15)	(0.16)	(0.17)	(0.18)	(0.18)
N	*60,000*	*60,000*	*60,000*	*60,000*	*60,000*	*60,000*
Mean of the control group	2.18	3.30	3.92	4.41	4.89	5.30

Notes: This table presents estimated effects of the individual and pooled treatments on whether students had connected at least once by each week of the six-week treatment period. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The sample in each regression includes all students participating in the study. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.2: Effects on Connection Status by Week (for Students that Connected At Least Once)

	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Individual treatments						
Streak	3.69*	13.43***	8.74***	9.48***	10.21***	10.59***
	(2.00)	(2.03)	(1.95)	(1.97)	(1.97)	(1.91)
Personalized Reminder (PR)	3.58*	7.96***	5.86***	6.49***	9.73***	7.86***
	(1.95)	(1.96)	(1.88)	(1.87)	(1.90)	(1.83)
Generic Reminder (GR)	3.20	3.97*	1.29	0.69	2.32	3.57*
	(2.16)	(2.14)	(2.05)	(2.01)	(2.02)	(1.97)
p-value (Streak = PR)	0.96	0.01	0.16	0.15	0.82	0.18
p-value (Streak = GR)	0.83	0.00	0.00	0.00	0.00	0.00
p-value (PR = GR)	0.86	0.07	0.03	0.01	0.00	0.04
Panel B: Pooled treatment						
Pooled treatment	3.51**	8.69***	5.55***	5.87***	7.81***	7.58***
	(1.61)	(1.59)	(1.52)	(1.52)	(1.52)	(1.47)
N	*4,146*	*4,146*	*4,146*	*4,146*	*4,146*	*4,146*
Mean of the control group	41.23	36.43	30.84	28.72	28.09	24.86

Notes: This table presents estimated effects of the individual and pooled treatments on whether students connect in each week of the six-week treatment period. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The sample in each regression includes the students who took theendline test. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.3: Effects on Having Streaks of Different Lengths - All Students

	Length of the streak				
	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks
	(1)	(2)	(3)	(4)	(5)
Panel A: Streaks of Weeks ***Connecting*** *to the Platform*					
Streak	1.87***	1.29***	0.97***	0.74***	0.56***
	(0.17)	(0.13)	(0.11)	(0.10)	(0.08)
Personalized Reminder (PR)	2.00***	1.23***	0.84***	0.48***	0.35***
	(0.17)	(0.13)	(0.11)	(0.09)	(0.07)
Generic Reminder (GR)	0.81***	0.34***	0.20**	0.05	0.06
	(0.14)	(0.11)	(0.09)	(0.07)	(0.06)
p-value (Streak = PR)	0.54	0.72	0.36	0.04	0.04
p-value (Streak = GR)	0.00	0.00	0.00	0.00	0.00
p-value (PR = GR)	0.00	0.00	0.00	0.00	0.00
N	*60,000*	*60,000*	*60,000*	*60,000*	*60,000*
Mean of the control group	1.53	0.83	0.53	0.39	0.25
Panel B: Streaks of Weeks ***not Connecting*** *to the Platform*					
Streak	-1.19***	-1.59***	-2.17***	-2.80***	-2.82***
	(0.13)	(0.16)	(0.20)	(0.24)	(0.26)
Personalized Reminder (PR)	-1.01***	-1.67***	-2.30***	-3.16***	-3.79***
	(0.12)	(0.16)	(0.20)	(0.24)	(0.27)
Generic Reminder (GR)	-0.14	-0.50***	-0.76***	-1.31***	-1.40***
	(0.09)	(0.13)	(0.18)	(0.22)	(0.24)
p-value (Streak = PR)	0.24	0.67	0.58	0.23	0.00
p-value (Streak = GR)	0.00	0.00	0.00	0.00	0.00
p-value (PR = GR)	0.00	0.00	0.00	0.00	0.00
N	*60,000*	*60,000*	*60,000*	*60,000*	*60,000*
Mean of the control group	99.29	98.63	97.39	96.05	94.70

Notes: This table presents estimated effects of the individual treatments on whether students maintain streaks of different lengths. Panel A reports the effects on streaks of connecting to the platform (i.e. consecutive weeks of connection). In contrast, Panel B reports effects on streaks of not connecting to the platform (i.e. consecutive weeks that students do not connect). Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate how long the streak is maintained. For example, the dependent variable for the regression reported in Panel A, column (1) corresponds to a dummy that equals one if the student had a two-week streak connecting to the platform during the six-week treatment period. The sample in each regression includes all students participating in the study. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.4: Effects on Having Streaks of Different Lengths - Students who Took the Baseline Test

	Length of the streak				
	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks
	(1)	(2)	(3)	(4)	(5)
Panel A: Streaks of Weeks ***Connecting*** *to the Platform*					
Streak	19.77***	15.91***	13.36***	11.63***	11.11***
	(2.71)	(2.47)	(2.24)	(2.07)	(1.93)
Personalized Reminder (PR)	21.81***	14.94***	10.59***	6.47***	5.91***
	(2.72)	(2.45)	(2.18)	(1.88)	(1.70)
Generic Reminder (GR)	8.91***	4.24*	1.92	0.52	-0.02
	(2.60)	(2.18)	(1.86)	(1.58)	(1.36)
p-value (Streak = PR)	0.54	0.75	0.32	0.04	0.03
p-value (Streak = GR)	0.00	0.00	0.00	0.00	0.00
p-value (PR = GR)	0.00	0.00	0.00	0.01	0.00
N	*2,268*	*2,268*	*2,268*	*2,268*	*2,268*
Mean of the control group	24.28	15.14	10.80	8.13	6.01
Panel B: Streaks of Weeks ***not Connecting*** *to the Platform*					
Streak	-15.93***	-17.94***	-17.77***	-17.46***	-14.54***
	(2.43)	(2.68)	(2.78)	(2.79)	(2.73)
Personalized Reminder (PR)	-12.71***	-16.18***	-18.17***	-18.24***	-17.91***
	(2.37)	(2.66)	(2.82)	(2.82)	(2.69)
Generic Reminder (GR)	-0.95	-4.54*	-3.24	-6.93**	-6.93**
	(2.03)	(2.50)	(2.72)	(2.83)	(2.82)
p-value (Streak = PR)	0.28	0.59	0.90	0.81	0.27
p-value (Streak = GR)	0.00	0.00	0.00	0.00	0.02
p-value (PR = GR)	0.00	0.00	0.00	0.00	0.00
N	*2,268*	*2,268*	*2,268*	*2,268*	*2,268*
Mean of the control group	85.75	77.17	66.70	57.57	47.66

Notes: This table presents estimated effects of the individual treatments on whether students maintain streaks of different length. Panel A reports the effects on streaks of connecting to the platform (i.e. consecutive weeks of connection). In contrast, Panel B reports the effects on streaks of not connecting to the platform (i.e. consecutive weeks that students do not connect). Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate how long the streak is maintained. For example, the dependent variable for the regression reported in Panel A, column (1) corresponds to a dummy that equals one if the student had a two-week streak connecting to the platform during the six-week treatment period. The sample in each regression includes the students who took the baseline test. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.5: Effects on Average Weekly Connection by Day of the Week

	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Individual treatments							
Streak	0.72***	0.16***	0.07***	0.42***	0.09***	0.03	0.06**
	(0.06)	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)
Personalized Reminder (PR)	0.59***	0.20***	0.08***	0.62***	0.09***	0.06***	0.02
	(0.05)	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)
Generic Reminder (GR)	0.45***	0.04	-0.01	0.03	-0.01	0.01	0.04*
	(0.05)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
p-value (Streak = PR)	0.09	0.28	0.82	0.00	0.98	0.36	0.12
p-value (Streak = GR)	0.00	0.00	0.01	0.00	0.00	0.33	0.60
p-value (PR = GR)	0.05	0.00	0.00	0.00	0.00	0.06	0.35
Panel B: Pooled treatment							
Pooled Treatment	0.59***	0.13***	0.05***	0.36***	0.06***	0.03**	0.04**
	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
N	*60,000*	*60,000*	*60,000*	*60,000*	*60,000*	*60,000*	*60,000*
Mean of the control group	0.39	0.30	0.25	0.23	0.22	0.15	0.15

Notes: This table presents estimated effects of the individual and pooled treatments on the percent of weeks connected by day of the week. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate the specific day on which the dependent variable is constructed. For example, the dependent variable in the regression reported in column (1) corresponds to the percent of Mondays in which the student connected during the six-week treatment period. The sample in each regression includes all students participating in the study. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.6: Heterogeneous Effects by Gender, Grade and Location on Connected at Least Once

	Gender		Grade			Location	
	Male	Female	Grade 4	Grade 5	Grade 6	Rural	Urban
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Individual treatments							
Streak	2.48***	3.25***	4.07***	2.85***	1.58***	2.30***	2.87***
	(0.33)	(0.40)	(0.49)	(0.46)	(0.38)	(0.60)	(0.28)
Personalized Reminder (PR)	3.54***	4.09***	4.62***	3.89***	2.74***	2.59***	3.96***
	(0.35)	(0.41)	(0.50)	(0.48)	(0.40)	(0.61)	(0.29)
Generic Reminder (GR)	1.41***	1.36***	2.44***	1.45***	0.33	1.07*	1.47***
	(0.32)	(0.37)	(0.47)	(0.43)	(0.34)	(0.55)	(0.27)
p-value (Streak = PR)	0.01	0.09	0.37	0.07	0.01	0.69	0.00
p-value (Streak = GR)	0.01	0.00	0.01	0.01	0.00	0.08	0.00
p-value (PR = GR)	0.00	0.00	0.00	0.00	0.00	0.03	0.00
Panel B: Pooled treatment							
Pooled treatment	2.48***	2.89***	3.71***	2.74***	1.54***	1.99***	2.76***
	(0.24)	(0.28)	(0.35)	(0.32)	(0.27)	(0.41)	(0.20)
N	*31,900*	*28,100*	*20,838*	*18,978*	*20,184*	*7,187*	*52,813*
Mean of the control group	4.70	5.96	7.08	5.33	3.44	2.75	5.64

Notes: This table presents estimated heterogeneous effects of the treatments on different measures of students' characteristics, considering gender, grade, and location. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate the sample included in the regression. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.7: Heterogeneous Effects by Baseline Use on Connected at Least Once

	Weekly connections in 2021		% correct exercises in 2021		Took the baseline test	
	Bottom 50	Top 50	Bottom 50	Top 50	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Individual treatments						
Streak	1.35***	4.57***	2.50***	3.13***	2.34***	14.54***
	(0.24)	(0.49)	(0.34)	(0.39)	(0.24)	(2.73)
Personalized Reminder (PR)	1.79***	6.15***	3.24***	4.33***	3.22***	17.91***
	(0.25)	(0.50)	(0.35)	(0.40)	(0.26)	(2.69)
Generic Reminder (GR)	0.82***	2.08***	1.22***	1.56***	1.18***	6.93**
	(0.22)	(0.46)	(0.32)	(0.37)	(0.23)	(2.82)
p-value (Streak = PR)	0.15	0.01	0.09	0.01	0.01	0.27
p-value (Streak = GR)	0.06	0.00	0.00	0.00	0.00	0.02
p-value (PR = GR)	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: Pooled treatment						
Pooled treatment	1.32***	4.27***	2.32***	3.01***	2.25***	13.13***
	(0.17)	(0.35)	(0.24)	(0.27)	(0.17)	(2.08)
N	*32 518*	*27,482*	*29,547*	*30,453*	*57,732*	*2,268*
Mean of the control group	2.03	9.16	4.13	6.41	3.47	52.34

Notes: This table presents estimated heterogeneous effects of the treatments on different measures of platform use. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate the sample included in the regression. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.8: Heterogeneous Effects by Gender, Grade and Location on Average Weekly Connection (for Students that Connected at Least Once)

	Gender		Grade			Location	
	Male	Female	Grade 4	Grade 5	Grade 6	Rural	Urban
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Individual treatments							
Streak	8.06***	10.76***	10.00***	9.58***	7.78***	5.19	9.67***
	(1.55)	(1.40)	(1.57)	(1.83)	(2.13)	(4.16)	(1.08)
Personalized Reminder (PR)	6.92***	7.36***	7.79***	8.43***	2.84	2.01	7.16***
	(1.39)	(1.30)	(1.50)	(1.63)	(1.76)	(3.37)	(0.99)
Generic Reminder (GR)	1.07	3.77***	2.45*	2.18	3.17	-2.81	2.81***
	(1.42)	(1.35)	(1.46)	(1.81)	(2.06)	(3.78)	(1.02)
p-value (Streak = PR)	0.48	0.03	0.19	0.55	0.02	0.45	0.03
p-value (Streak = GR)	0.00	0.02	0.00	0.00	0.88	0.20	0.00
p-value (PR = GR)	0.00	0.00	0.00	0.00	0.06	0.07	0.00
Panel B: Pooled treatment							
Pooled treatment	5.74***	7.45***	6.95***	7.14***	4.59***	1.63	6.78***
	(1.14)	(1.03)	(1.18)	(1.33)	(1.52)	(2.90)	(0.79)
N	*1,971*	*2,175*	*1,928*	*1,323*	*895*	*291*	*3,855*
Mean of the control group	32.61	30.88	33.65	31.40	28.00	32.05	31.67

Notes: This table presents estimated heterogeneous effects of the treatments on different measures of students' characteristics, considering gender, grade, and location. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate the sample included in the regression. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.9: Heterogeneous Effects by Baseline Use on Average Weekly Connection (for Students that Connected at Least Once)

	Weekly connections in 2021		% correct exercises in 2021		Took the baseline test	
	Bottom 50	Top 50	Bottom 50	Top 50	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Individual treatments						
Streak	9.54***	9.33***	9.62***	9.31***	6.56***	14.91***
	(2.03)	(1.20)	(1.47)	(1.41)	(1.05)	(2.28)
Personalized Reminder (PR)	4.77***	7.46***	7.42***	6.61***	5.16***	10.27***
	(1.67)	(1.12)	(1.35)	(1.28)	(0.95)	(2.15)
Generic Reminder (GR)	1.89	2.69**	4.26***	1.47	2.54***	1.92
	(1.69)	(1.16)	(1.40)	(1.34)	(0.98)	(2.16)
p-value (Streak = PR)	0.02	0.15	0.17	0.07	0.23	0.06
p-value (Streak = GR)	0.10	0.00	0.04	0.00	0.02	0.00
p-value (PR = GR)	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: Pooled treatment						
Pooled treatment	5.52***	6.76***	7.29***	6.06***	4.92***	9.35***
	(1.40)	(0.89)	(1.07)	(1.04)	(0.75)	(1.68)
N	*922*	*3,224*	*1,668*	*2,478*	*2,778*	*1,368*
Mean of the control group	25.95	33.20	27.09	34.54	24.99	43.12

Notes: This table presents estimated heterogeneous effects of the treatments on different measures of platform use. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate the sample included in the regression. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.10: Effects on Answered any Exercise and % of Attempted Exercises in the Endline Test

	Answered any question	% of attempted questions in the endline test (for students taking the test)
	(1)	(2)
Panel A: Individual treatments		
Streak	0.71***	7.71***
	(0.17)	(2.61)
Personalized Reminder (PR)	0.38**	3.02
	(0.17)	(2.75)
Generic Reminder (GR)	-0.05	3.78
	(0.16)	(2.84)
p-value (Streak = PR)	0.10	0.11
p-value (Streak = GR)	0.00	0.20
p-value (PR = GR)	0.02	0.81
Panel B: Pooled treatment		
Pooled treatment	0.35***	5.03**
	(0.12)	(2.09)
N	*60,000*	*1,503*
Mean of the control group	2.29	66.13

Notes: This table presents estimated effects of the individual and pooled treatments on two outcomes related to the endline test. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate the dependent variable in the regression. The dependent variable in column (1) is an indicator that equals one if the student attempted at least one question in the endline test (i.e. took the endline test). The dependent variable in column (2) corresponds to the percent of questions (out of 30) that the student attempted in the endline test. The sample for the regression reported in column (1) includes all students participating in the study. The sample for the regression reported in column (2) includes students who took theendline test. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.11: Effects on Platform Use - Students Who Took the Endline Test

	Connected at least once (in %)	Average weekly connection (for students that connected at least once, %)	Average weekly connection (for all students, %)	Average exercises attempted (%)
	(1)	(2)	(3)	(4)
Panel A: Individual treatments				
Streak	22.10***	15.28***	22.49***	20.97***
	(2.77)	(2.30)	(2.20)	(2.22)
Personalized Reminder (PR)	23.69***	12.93***	21.13***	17.51***
	(2.84)	(2.35)	(2.19)	(2.19)
Generic Reminder (GR)	9.01***	4.13	6.52***	5.74**
	(3.35)	(2.62)	(2.33)	(2.27)
p-value (Streak = PR)	0.58	0.34	0.59	0.18
p-value (Streak = GR)	0.00	0.00	0.00	0.00
p-value (PR = GR)	0.00	0.00	0.00	0.00
Panel B: Pooled treatment				
Pooled treatment	18.91***	11.70***	17.50***	15.48***
	(2.34)	(1.89)	(1.68)	(1.66)
N	*1,503*	*1 052*	*1,503*	*1,503*
Mean of the control group	58.36	51.77	30.21	25.98

Notes: This table presents estimated effects of the individual and pooled treatments on different measures of platform use. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate the dependent variable in the regression. The dependent variable in column (1) is an indicator that equals one if the student connected at least once during the six-week treatment period. The dependent variable in column (2) and (3) corresponds to the percent of weeks that the student connected during the six-week treatment period. The dependent variable in column (4) corresponds to the percent of exercises attempted out of all the 180 exercises provided to students during the six-week treatment period. The sample in the regressions reported in columns (1), (3) and (4) includes the students who took the endline test. The sample in the regression reported in column (2) includes only students who took the endline test and who connected at least once during the six-week treatment period. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.12: Effect of Usage on Math Achievement

	(1)	(2)	(3)	(4)
	Panel A: Individual treatments as IV			
Proportion of weeks connected (for all students)	0.48**	0.57***	0.35	0.43**
	(0.23)	(0.20)	(0.23)	(0.20)
First-Stage F	50.48	54.08	50.48	52.56
	Panel B: Pooled treatment as IV			
Proportion of weeks connected (for all students)	0.54**	0.66***	0.59**	0.66***
	(0.27)	(0.24)	(0.28)	(0.25)
First-Stage F	108.04	116.24	108.04	112.02
N	*1,503*	*1,503*	*1,503*	*1,503*
Questions in test included	All	All	Attempted	Attempted
2021 controls	N	Y	N	Y

Notes: This table presents estimated effects of the proportion of weeks connected on math achievement. Each column in a panel corresponds to a separate 2SLS regression. Labels in rows correspond to independent variables. Math achievement in columns (1) and (2) is computed considering all 30 questions included in the endline test. Math achievement in columns (3) and (4) is computed considering only the questions attempted by the student (i.e. questions not answered are not included to compute the percent of correct answers). The sample in each regression includes the students who took the endline test. All regressions include strata fixed-effects. Regressions reported in columns (2) and (4) also control for average math performance in 2021. Math achievement has been normalized by subtracting the mean and dividing by the standard deviation of the control group. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.13: Effects on Platform Use (for Students that Took the Endline Test)

	Connected at least once (%)	Average weekly connection (for students that connected at least once, %)	Average weekly connection (for all students, %)	Average exercises attempted (%)
	(1)	(2)	(3)	(4)
Panel A: Individual treatments				
Streak	22.10***	15.28***	22.49***	20.97***
	(2.77)	(2.30)	(2.20)	(2.22)
Personalized Reminder (PR)	23.69***	12.93***	21.13***	17.51***
	(2.84)	(2.35)	(2.19)	(2.19)
Generic Reminder (GR)	9.01***	4.13	6.52***	5.74**
	(3.35)	(2.62)	(2.33)	(2.27)
p-value (Streak = PR)	0.58	0.34	0.59	0.18
p-value (Streak = GR)	0.00	0.00	0.00	0.00
p-value (PR = GR)	0.00	0.00	0.00	0.00
Panel B: Pooled treatment				
Pooled treatment	18.91***	11.70***	17.50***	15.48***
	(2.34)	(1.89)	(1.68)	(1.66)
N	*1,503*	*1,052*	*1,503*	*1,503*
Mean of the control group	58.36	51.77	30.21	25.98

Notes: This table presents estimated effects of the individual and pooled treatments on different measures of platform use. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate the dependent variable in the regression. The dependent variable in column (1) is an indicator that equals one if the student connected at least once during the six-week treatment period. The dependent variable in column (2) and (3) corresponds to the percent of weeks that the student connected during the six-week treatment period. The dependent variable in column (4) corresponds to the percent of exercises attempted out of all the 180 exercises provided to students during the six-week treatment period. The sample in the regressions reported in columns (1), (3) and (4) includes all students. The sample in the regression reported in column (2) includes only students who connected at least once during the six-week treatment period. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.14: Effects of learning platforms on academic achievement in developing countries

Study	Country	Effects on Math	Effects on Language	Target Subject
Araya et al. (2025)	Chile	0.27	-0.01	Math
Bai et al. (2016)	China	-	0.06	Language
Bai et al. (2018)	China	-	0.56	Language
Bettinger et al. (2020)	Russia	0.11	0.03	Both
Büchel et al. (2022)	El Salvador	0.23	-	Math
Cardim et al. (2023)	Angola	-0.01	0.00	Both
Fafchamps and Mo (2018)	China	0.17	-	Math
Ferman, Finamor, and Lima (2019)	Brazil	-0.02	-	Math
Ito et al. (2019)	Cambodia	0.73	-	Math
Lai et al. (2012)	China	0.22	0.19	Both
Lai et al. (2013)	China	0.19	-	Math
Lai et al. (2015)	China	0.21	0.11	Math
Lai et al. (2016)	China	0.11	0.15	Both
Levesque et al. (2022)	Malawi	0.54	0.37	Both
Ma et al. (2020)	China	0.03	-	Math
Mo et al. (2014)	China	0.16	0.05	Math
Mo et al. (2020)	China	-	0.05	Language
Muralidharan et al. (2019)	India	0.37	0.23	Both

Notes: This table shows a summary of the evidence on the effects of learning platforms on academic achievement in developing countries. The following criteria were used to include studies: (i) implemented in primary or secondary school; (ii) aimed to improve learning in math or language; (iii) results are reported in papers or other sources published after 2011; (iv) compared the treatment group to the status quo; (v) effects measured at least 12 weeks after the start of the program; (vi) effects estimated using experimental evaluations; (vii) samples included at least 200 students; (viii) at least 20 units (classes or schools) were randomized. Effects are expressed in standard deviations. Intent-to-treat (ITT) estimated effects are reported in the table.

Table A.15: Exploring External Validity - Characteristics of Relevant School Samples

	All	Public	Public and in five regions focused by Conecta Ideas Peru	With students who participated in the study
	(1)	(2)	(3)	(4)
		School characteristics		
Urban (%)	74.16	67.22	85.34	88.07
Public (%)	78.17	100.00	100.00	100.00
Coastal region (%)	47.13	40.42	57.00	51.77
Andean region (%)	34.22	36.63	36.24	44.08
Jungle region (%)	18.65	22.95	6.76	4.15
		School access to services		
Drinking water (%)	78.80	74.90	89.13	90.41
Sewage (%)	77.05	72.74	88.09	89.38
Electricity (%)	90.68	89.66	96.24	96.82
Internet (%)	60.31	54.79	68.54	68.03
		Observations		
Schools	*38,523*	*29,903*	*6,288*	*1,820*
Students in 4th to 6th grade	*1,888,566*	*1,476,263*	*536,871*	*60,000*

Notes: This table presents statistics for different samples of students in 4th to 6th grade in Peruvian schools. Column (1) presents means for the overall population of students in 4th to 6th grade in all schools in Peru. Column (2) restricts the sample from column (1) to public schools, and column (3) focuses on public schools in the five regions where Conecta Ideas had at least 2% of the Conecta Ideas student sample: Arequipa, Cusco, Ica, Junin and Lima. For columns (1) to (3), statistics are weighted by the number of students in each school from 4th to 6th grades. Column (4) considers only the students in the schools where there is at least one student from the sample of 60,000 students participating in the study and weights school observations by the number of students in the study sample per school.

Table A.16: Heterogeneous Effects by Gender, Grade and Location on Average Weekly Connection

	Gender		Grade			Location	
	Male	Female	Grade 4	Grade 5	Grade 6	Rural	Urban
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Individual treatments							
Streak	1.31***	1.85***	2.30***	1.52***	0.83***	0.94***	1.62***
	(0.15)	(0.18)	(0.23)	(0.21)	(0.16)	(0.25)	(0.13)
Personalized Reminder (PR)	1.55***	1.80***	2.19***	1.83***	0.90***	0.90***	1.75***
	(0.15)	(0.17)	(0.23)	(0.21)	(0.14)	(0.23)	(0.13)
Generic Reminder (GR)	0.48***	0.63***	0.90***	0.57***	0.24**	0.24	0.61***
	(0.12)	(0.15)	(0.19)	(0.17)	(0.12)	(0.20)	(0.11)
p-value (Streak = PR)	0.00	0.00	0.00	0.00	0.00	0.01	0.00
p-value (Streak = GR)	0.20	0.80	0.69	0.25	0.71	0.89	0.43
p-value (PR = GR)	0.00	0.00	0.00	0.00	0.00	0.01	0.00
Panel B: Pooled treatment							
Pooled treatment	1.12***	1.42***	1.80***	1.31***	0.66***	0.70***	1.33***
	(0.10)	(0.12)	(0.15)	(0.13)	(0.10)	(0.16)	(0.08)
N	*31,900*	*28,100*	*20,838*	*18,978*	*20,184*	*7,187*	*52,813*
Mean of the control group	1.53	1.84	2.38	1.67	0.96	0.88	1.79

Notes: This table presents estimated heterogeneous effects of the treatments on different measures of students' characteristics, considering gender, grade, and location. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate the sample included in the regression. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.17: Heterogeneous Effects by Baseline Use on Average Weekly Connection

	Weekly connections in 2021		% correct exercises in 2021		Took the baseline test	
	Bottom 50	Top 50	Bottom 50	Top 50	No	Yes
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Individual treatments						
Streak	0.68***	2.59***	1.27***	1.82***	0.97***	16.22***
	(0.10)	(0.23)	(0.14)	(0.19)	(0.09)	(2.01)
Personalized Reminder (PR)	0.62***	2.88***	1.36***	1.94***	1.13***	14.87***
	(0.09)	(0.23)	(0.14)	(0.18)	(0.09)	(1.91)
Generic Reminder (GR)	0.27***	0.89***	0.54***	0.56***	0.41***	4.10**
	(0.07)	(0.19)	(0.11)	(0.15)	(0.07)	(1.73)
p-value (Streak = PR)	0.00	0.00	0.00	0.00	0.00	0.00
p-value (Streak = GR)	0.65	0.33	0.61	0.60	0.17	0.57
p-value (PR = GR)	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: Pooled treatment						
Pooled treatment	0.53***	2.12***	1.05***	1.44***	0.84***	11.74***
	(0.06)	(0.15)	(0.09)	(0.12)	(0.06)	(1.36)
N	*32,518*	*27,482*	*29,595*	*30,405*	*57,732*	*2,268*
Mean of the control group	0.53	3.04	1.11	2.22	0.87	22.57

Notes: This table presents estimated heterogeneous effects of the treatments on different measures of platform use. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. The column titles indicate the sample included in the regression. All regressions include strata fixed-effects. Robust standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, ** and *, respectively.

Table A.18: Difference in Baseline Characteristics Based on Students' Participation in the Endline Test

	Took the endline Test (1)	Did not take the endline test (2)	Difference (3)	*N* (4)
School characteristics				
Multigrade	3.66	6.07	-2.41	*60,000*
			(0.62)***	
Located in Lima	50.57	49.13	1.43	*60,000*
			(1.31)	
Student characteristics				
Female	55.14	50.70	4.44	*55,302*
			(1.35)***	
Fourth grade	49.43	34.35	15.08	*60,000*
			(1.24)***	
Fifth grade	32.07	31.62	0.45	*60,000*
			(1.21)	
Weeks connected in 2021	10.06	5.35	4.70	*60,000*
			(0.12)***	
Math achievement in 2021	0.33	-0.02	0.35	*59,093*
			(0.03)***	

Notes: This table presents statistics and estimated differences between students that took and did not take the endline test. The table reports statistics for all students participating in the study. Column (1) presents means for the sample of students who took the endline test. Column (2) presents means for the sample of students who did not take the endline test. Column (3) presents the estimated coefficients and standard errors for differences in characteristics between students that took and those that did not take the endline test. Standard errors are reported in parentheses. Significance at the one, five and ten percent levels is indicated by ***, **, and *, respectively.

Appendix A.1: Messages included in app notifications

This appendix presents the messages used in the baseline and endline weeks and those sent during the six-week treatment period. Two versions of each message were sent during the treatment period, alternating between even and odd weeks. Both types will be presented here, annotated by (O) for odd and (E) for even. The original messages, sent in Spanish, are presented below alongside their English translations.

1. **Initial message informing of the summer program and promoting participation in the baseline test**

All 60,000 students participating in the experiment received the following message on Monday, January 10, 2022 during the baseline test week:

"Would you like your child to continue practicing mathematics during the summer? 🌞 Encourage him/her to log into the Conecta Ideas app and keep earning flags🚩. If you do not wish to receive these messages, please write a message to WhatsApp at +51 991 726 718."

Also, the following message was sent to all students on Thursday, January 13, 2022:

"👋 Would you like to practice math today with Conecta Ideas? Log in and keep earning flags! "

Original messages in Spanish:

"¿Quisiera que su hijo/a siga practicando matemática durante el verano 🌞? Motívelo/a a que ingrese al app de Conecta Ideas y siga acumulando banderitas🚩. Si no quiere recibir estos mensajes, escriba al WhatsApp +51 991 726 718"

"👋 ¿Te animas a practicar matemática hoy con Conecta Ideas? ¡Ingresa y sigue ganando banderitas! "

2. Simple Reminder

During the six-week treatment period, all three treatment groups received a message every Monday reminding them that a set of math exercises was available to complete that week. The message stated: "We have a new activity on Conecta Ideas! Join, practice, and earn more flags! 🚩" .

In Spanish: "¡Ya tenemos una nueva actividad en Conecta Ideas! ¡Ingresa, practica y gana más banderitas! 🚩"

3. Personalized Reminder

Students in the Personalized Reminder group not only received the simple reminder as presented above but also a personalized reminder every Thursday during the six-week treatment period that varied depending on whether students had already connected that week. During odd weeks (1, 3 and 5), the following messages were sent:

Connected this week	Message
Yes	Congratulations, Ana! 🎉 This week, you participated for more flags. See you again on Monday!
No	Hello, Ana! ⏰ There's still time to practice mathematics. Participate for more flags!

During even weeks (2, 4 and 6), the following messages were sent:

Connected this week	**Message**
Yes	Well done, Ana! This week, you did math exercises. We will see you on Monday!
No	Hello, Ana! Today is a great day to practice mathematics. Participate and earn more flags!

The original messages in Spanish for odd weeks are presented next:

Connected this week	**Message**
Yes	¡Felicitaciones, Ana! Esta semana participaste por más banderitas. ¡Nos vemos de nuevo el lunes!
No	¡Hola, Ana! Aún hay tiempo para practicar matemática. ¡Participa por más banderitas!

The original messages in Spanish for even weeks are presented next:

Connected this week	**Message**
Yes	¡Bravo, Ana! Esta semana hiciste ejercicios de matemática. ¡Te esperamos el lunes!
No	¡Hola, Ana! Hoy es un gran día para practicar matemática. ¡Participa y gana más banderitas!

4. Streaks

Students in the Streak treatment group received not only the simple reminder, as presented above, but also additional messages that varied depending on their engagement in the current and preceding weeks. The messages in English, for odd weeks, are presented next:

<table>
<tr><td colspan="2" rowspan="2"></td><td colspan="2">Connected the previous week</td></tr>
<tr><td>Yes</td><td>No</td></tr>
<tr><td rowspan="2">Connected this week</td><td>Yes</td><td>Great job Ana! 🎉 You've accumulated 3 consecutive weeks of practicing mathematics. Let's continue this Monday for more!</td><td>Welcome, Ana! 🎊 You've started accumulating weeks in Conecta Ideas. Let's continue this Monday for more!</td></tr>
<tr><td>No</td><td>Hello Ana! 👏 Today, you could reach 3 consecutive weeks of practicing mathematics. We're waiting for you!</td><td>Hello Ana! 👏 Today, you could start accumulating weeks of practice in Conecta Ideas. We're waiting for you!</td></tr>
</table>

The messages in Spanish, for odd weeks, are presented next:

		Connected the previous week	
		Yes	No
Connected this week	Yes	¡Bien hecho Ana! 🎉 Acumulaste 3 semanas consecutivas practicando matemática. ¡Sigamos el lunes por más!	¡Bienvenida Ana! 🎊 Comenzaste a acumular semanas en Conecta Ideas. ¡Sigamos el lunes por más!
	No	¡Hola Ana! 👏 Hoy podrías alcanzar 3 semanas consecutivas practicando matemática. ¡Te esperamos!	¡Hola Ana! 👏 Hoy podrías comenzar a acumular semanas de práctica en Conecta Ideas. ¡Te esperamos!

The messages in English, for even weeks, are presented next:

		Connected the previous week	
		Yes	No
Connected this week	Yes	Well done Ana! 🎉 You've been practicing mathematics for 3 consecutive weeks. See you on Monday!	Great to have you back, Ana! 🎊 You've accumulated 1 week of practice in Conecta Ideas. See you on Monday!
	No	👏 Ana, today you would have 3 consecutive weeks of practicing mathematics. Don't miss this achievement. We encourage you to participate!	👏 Ana, today is a good time to accumulate weeks of practice in Conecta Ideas. We encourage you to participate!

The messages in Spanish, for even weeks, are presented next:

		Connected the previous week	
		Yes	No
Connected this week	Yes	¡Bravo Ana! 🎉 Llevas 3 semanas seguidas practicando matemática. ¡Nos vemos el lunes!	¡Qué bueno tenerte de vuelta Ana! 🪅 Acumulaste 1 semana de práctica en Conecta Ideas. ¡Nos vemos el lunes!
	No	👋 Ana, hoy tendrías 3 semanas seguidas practicando matemática. No pierdas este logro. ¡Animate a participar!	👋 Ana, hoy es un buen momento para acumular semanas de práctica en Conecta Ideas. ¡Animate a participar!

5. Message to promote participation in the endline test

During the endline test week, all 60,000 students participating in the experiment received the following message on Monday, February 28, 2022:

"Come and end the summer with Conecta Ideas! 🌞 We have a new review activity available."

And on Thursday, March 3, 2022, all students received the following message:

"👋 Would you like to practice math today with Conecta Ideas? Log in and keep earning flags!"

The messages in Spanish are the following:

"¡Ven y cierra el verano con Conecta Ideas! 🌞 Ya tenemos disponible una nueva actividad de repaso."

"👋 ¿Te animas a practicar matemática hoy con Conecta Ideas? ¡Ingresa y sigue ganando banderitas!"

Appendix A.2 – External validity

This appendix explores the external validity of our main findings. The sample used to measure effects on platform use includes 60,000 students in schools that participated in the Conecta Ideas Peru program in 2021 and that used the learning app in that year. These students attended schools that voluntarily signed up to use the Conecta Ideas learning app and belonged to families with access to smartphones and the willingness to use them to practice math. This raises the question of the representativeness of the study sample compared to the general student population of Peru.[18]

We begin with Appendix Table A.15 which documents the school selection process and shows that, while schools located in the regions that implemented Conecta Ideas Peru tended to be more urban and had better access to services compared to all public schools in the country, the schools with students that participated in the experiment were quite similar to the sample of schools in the regions focused by Conecta Ideas Peru. Unfortunately, we do not have student-level data to assess student selection within schools. However, we might expect that students with higher access to technology and with parents more supportive of their education were more likely to be included in the study. Moreover, since we document substantial heterogeneity in the effects across students with different observable characteristics, a question emerges about how the effects could differ if the study sample were more similar to the general student population.

Though we cannot provide a definite answer to this question, appendix tables A.6 to A.9, discussed briefly in Section 4.1, suggest that the main findings of this study are likely to extend more broadly. Appendix tables A.6 and A.7 document that the estimated effects on the extensive

[18] Like many developing nations, Peru has experienced a rapid increase in access to smartphones. Data from the 2015 OECD's Programme for International Student Assessment (PISA) shows that 73% of 15-year-olds in Peru reported having a smartphone at home, a figure that rose to 81% by 2018. In comparison, the average for developing countries participating in PISA was higher, starting at 89% in 2015 and reaching 95% by 2018.

margin are larger for the Personalized Reminder treatment compared to the Streaks treatment in all 13 subsamples of students defined by gender, grade, location, baseline connection, baseline academic achievement, and participation in the baseline test (in 9 cases the difference is statistically significant at the 10 percent level). Similarly, Appendix tables A.8 and A.9 show that in all 13 subsamples, the estimated effects on the intensive margin of the Streaks treatment are larger than the Personalized Reminder treatment (in 5 cases the difference is statistically significant at the 10 percent level). Finally, we checked that in none of the 13 subsamples the overall effects on percent of weeks connected are statistically significantly different between the Personalized Reminder treatment and the Streaks treatment (tables A.16 and A.17).

Regarding the external validity of the effects on academic achievement, we explore whether the students that participated in the endline test (N=1,503) have similar characteristics compared to students that did not participate in this test (N=58,407). Appendix Table A.18 documents that students participating in the endline test had a lower likelihood of attending a multigrade school (i.e. a rural school in which a teacher covers more than one grade), were more likely to be female, to be in grade 4 and to have higher baseline engagement with the platform, and to have higher baseline academic performance. Since the sample of students participating in the endline test is limited, we do not have sufficient statistical power to replicate the analysis exploring the robustness of the results on learning when focusing on specific subsamples as we did for the results on platform use. Consequently, these results are more tentative and future research could provide further evidence on whether highlighting streaks can produce increases in academic achievement for other populations.