WORKING PAPER Nº IDB-WP-1566

Streaking to Success:

The Effects of Highlighting Streaks on Student Effort and Learning

Raphaëlle Aulagnon
Julián Cristia
Santiago Cueto
Ofer Malamud

Inter-American Development Bank
Department of Research and Chief Economist

April 2024

Streaking to Success:

The Effects of Highlighting Streaks on Student Effort and Learning

Raphaëlle Aulagnon*
Julián Cristia**
Santiago Cueto***
Ofer Malamud****

* Bocconi University
** Inter-American Development Bank
*** GRADE
**** Northwestern University

Inter-American Development Bank
Department of Research and Chief Economist

April 2024

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Streaking to success: the effects of highlighting streaks on student effort and learning / Raphaëlle Aulagnon, Julian Cristia, Santiago Cueto, Ofer Malamud.
p. cm. - (IDB Working Paper Series ; 1566)
Includes bibliographical references.
1. Fiscal policy-Central America. 2. Equality-Central America. 3. Gender mainstreaming-Central America. 4. Wages-Women-Central America. 5. Equal pay for equal work-Central America. I. Aulagnon, Raphaëlle. II. Cristia, Julia. III. Cueto, Santiago, 1960- IV. Malamud, Ofer. V. Inter-American Development Bank. Department of Research and Chief Economist. VI. Series.
IDB-WP-1566

http://www.iadb.org

Abstract[*]

We examine whether highlighting streaks encourages 4th to 6th grade students in Peru to increase their use of an online math platform and improve learning. Sixty thousand students were randomly assigned to receive messages that i) highlighted streaks, ii) provided personalized reminders with positive reinforcement, or iii) provided generic reminders, while others were assigned 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.

JEL classifications: I21, I25, D91
Keywords: Streaks, Education, Math achievement, Online

* 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 that included, among others, Carlos Aguirre, Raúl Gormaz and Paulina Jaure. A team at the research center GRADE that was coordinated by Natalia Arteta, Jessica Simon and Claudia Sugimaru and included Giulia Celi, Sergio García, Giancarlo Garrido, Antonio Iglesias, Lucciano Ognio, Ingrid Olea, Roberto Palacios, Flor Paredes, André Sánchez and Jose Segovia was responsible for the scale up of the learning platform used for this study. 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 used in this study. 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 (Comité Institutional de Bioética) ethics board and registered the Conecta Ideas data with the Ministry of Justice in Peru.

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 and completing assignments on time (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 by 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 from the behavior on which it is based (Silverman and Barasch, 2023). This becomes a source of motivation as individuals try to avoid losing the streak which may be further amplified by loss aversion. 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 widespread 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, which tracks and informs users about the number of days in a row that they have completed a lesson. 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

[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. The use of streaks is also pervasive on social networking sites like Snapchat and fitness apps like Fitbit.

retire its streak feature due to worries that it might negatively impact motivation in certain circumstances.[3] However, despite the pervasive use of streaks and debates about their possible effects, there is limited empirical evidence examining their causal impacts.

Our study examines 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. We conducted a randomized field experiment during the summer of 2022 that examined platform use and learning outcomes for students in households that were sent messages which i) highlighted streaks of completed assignments, as compared to those which ii) provided personalized reminders with positive reinforcement, or iii) provided generic reminders; others were assigned to a control group which did not receive any messages.[4] The online math platform was aligned with the national math curriculum and provided a weekly set of 30 math exercises. The experiment was targeted at 60,000 students in households that had downloaded and used the math platform during the preceding academic year. Messages were sent through app notifications over the six weeks of treatment, and a baseline test and an endline test were administered through the app in the weeks before and after the treatment period.

Perhaps not surprisingly, since it was not required by teachers or schools, 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 31.7 percent of weeks in the treatment period. Yet 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). However, highlighting streaks showed larger effects than personalized reminders on the intensive margin, in terms of fraction of weeks connected conditional on connecting at least once (9.4 vs. 6.9 percentage points).[5] The effects on these two margins counterbalanced each other, leading to similar effects of highlighting streaks and providing

[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] Section 2.3 and Appendix A.1 describes the treatments in detail.
[5] The effect of providing generic reminders on use was also significantly larger than the control group, but smaller than those of highlighting streaks and providing personalized reminders (on both the extensive and intensive margin).

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.

We do find evidence of significant effects on learning. Pooling our treatments, we observe a 0.10 to 0.12 standard deviation increase in performance on the endline math test that was administered in the week immediately after the six-week treatment period. However, this effect is driven mostly by the notifications of streaks, the only treatment arm for which the effects are consistently large and significant; students whose streaks were highlighted scored approximately 0.13 to 0.17 standard deviations higher than the control group. 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. The only other study examining the effects of highlighting 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 serve to increase subsequent engagement relative to when "broken" streaks are highlighted. Our study examines the role of highlighting streaks in a nationwide field experiment with an online math curriculum supported by the Peruvian Ministry of Education and 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. (2019) and Muralidharan et al. (2019) show 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 Rodríguez-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 only 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 percent to 23 percent, 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," was developed by a team at the Center for Advanced Research on Education at the Universidad de Chile. After years of small-scale development, the platform was used in 11 primary schools in Santiago, Chile, between 2011 and 2016. 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., 2019). In 2018-2019, the Peruvian research center GRADE adapted the platform for use in schools in Lima, Peru with the support of the Universidad de Chile, the Ministry of Education of Peru, the International Development Research Center, and the Inter-American Development Bank. GRADE further modified the program in 2020-2021 to support home-based learning since most public schools in Peru were closed during the pandemic.

To use the platform, teachers are required to register themselves and their students by completing a form on the Conecta Ideas Peru website. Teachers need to share student usernames with parents so that they can log on to the platform. Parents can then download the app and log in with the information received from teachers so that their children can use the learning app.[6]

The Conecta Ideas app provides students with 30 weekly math exercises that they can complete using a smartphone, tablet, or computer (the vast majority use a smartphone). The platform is free, and students only need 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 have

[6] After installing the app, parents can choose whether they wish to receive notifications as with other apps.

logged on, students are presented with 30 math exercises that they need to solve within the week. After solving each exercise, they receive auditory and visual feedback letting them know whether their answer was correct or not. Students are awarded "flags" for solving exercises correctly and those who complete the 30 exercises are presented with a summary of the "flags" they have accumulated in that week, month, and during the year.

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.[7] 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 the app were presented with a set of 30 math exercises among four broad topics: i) numeracy, ii) geometry, iii) probability, and iv) patterns.[8] 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 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.[9] Two app notifications were also sent to all households on 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 control group did not receive any further notifications, while the three treatment groups received different types of app notifications each week.

[7] The 2021 academic year ended on December 14, 2021, and the 2022 academic year started on March 14, 2022.
[8] Weeks 2, 3 and 6 covered numeracy, while weeks 1, 4 and 5, covered geometry, probability, and patterns.
[9] Appendix A.1 presents the messages sent during the experiment for students in the different treatment arms.

The three different types of notifications were: 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 p.m. 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 p.m. 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, as described in Appendix A.1. These messages 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 p.m. 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 a positive reinforcement message to those who had completed it.

It is important to note that 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

We exploit detailed data from the Conecta Ideas platform for our analysis. The primary dataset contains information at the student-exercise level about when each exercise was completed and whether these responses are correct. A complementary file contains the student internal ID and her/his grade. Student gender was not recorded in the platform but assigned for 93 percent 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. Participation in the baseline and endline tests was low: only 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

We exploit 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., number of weeks connected or average percent of correct responses in 2021).

We also provide evidence regarding how platform use affected math achievement by exploiting variation in which topics students were exposed to during the 6-week treatment period. Specifically, we compare students who connected to the math platform on weeks when certain

topics (e.g., numeracy, geometry) were covered to students who connected on weeks that covered other topics. We implement this strategy by first estimating the following equation:

$$Y_i = \beta_0 + \beta_1 Numeracy_i + \beta_2 Geometry_i + \beta_3 Probability_i + \beta_4 Patterns_i + \eta_w + \tau_i \quad (2)$$

where Y_i is the standardized achievement for a given topic (e.g. numeracy) on the endline test for individual i. The variables *Numeracy*, *Geometry*, *Probability*, and *Patterns* are measures of exposure to these topics during the treatment period.[10] The coefficients $\beta_1, \beta_2, \beta_3,$ and β_4 represent corresponding estimates for the effect of exposure to these topics relative to students who did not connect at all. In addition, η_w are fixed effects for the number of weeks that student i used the platform during the treatment period, and τ_i is the individual-level error term.

We summarize these effects by estimating a stacked version of the regression above where we regress the standardized endline score for a particular topic on an indicator for whether students were exposed to this topic during the treatment period.

$$Y_{ij} = \alpha_0 + \alpha_1 Exposed_{ij} + \eta_w + \mu_j + \tau_{ij} \quad (3)$$

where Y_{ij} is standardized achievement in topic j on the endline test for individual i, $Exposed_{ij}$ is an indicator for 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.[11] The coefficient α_1 represents the effect of being exposed to the math platform on weeks when a tested topic (e.g., geometry) was covered relative to students who connected on weeks that covered other topics. 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

[10] The variables *Geometry*, *Probability* and *Patterns* are indicators for having used the platform during the week that each topic was covered in the platform. In the case of *Numeracy*, because this topic was covered during three weeks in the training period, the corresponding variable is equal to the fraction of weeks that the student used the platform when this topic was covered.

[11] We also present results for alternative specifications that include topic fixed-effects and that replace the dummies for number of weeks connected with student fixed-effects and topic fixed-effects.

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 statistically significant, except for percent of correct exercises 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 who 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

Table 2 reports treatment effects on: 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 overall average weekly connection; and iv) the fraction of exercises attempted.[12] Since teachers and schools did not require the use of the math learning platform over the summer, take-up was relatively low: only 5.3 percent of 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 attempted, without accounting for the large number of students who did not connect, was only 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

[12] The average weekly connection corresponds to the fraction of weeks that students connect during the treatment period.

[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.

relative to the control group, which is lower than the effect of sending personalized reminders (3.8pp) but higher than the effect of sending generic reminders (1.4pp).[14] 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 percentage point 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 (6.9pp) and sending generic reminders (2.5pp). 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 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 attempted, 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 Reminder but similar to those from Personalized Reminder. But did the Streak treatment actually produce streaks (i.e., repeated connections in consecutive weeks)? To explore this, Appendix Table A.3 presents the effects of the treatments on the fraction of students who used the platform consecutively for two, three, and 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

[14] All effects discussed in the paper are statistically significant effects at the 5 percent level unless noted otherwise.

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.[15]

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 in favor of this hypothesis (recall that both treatments had similar effects in overall engagement measured by percent of weeks connected or exercises attempted).

We also analyze whether the effects are fully 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 indeed concentrated on days when messages are sent (i.e., Monday for the three treatments and Thursdays for Streak and Personalized Reminder). Still, there are also increased connections during other days of the week for the Streak and Personalized Reminder treatments.

We explore heterogeneous effects on the extensive margin (i.e. connecting at least once) in Appendix Tables A.6 and A.7. We find that the impact of the Streak treatment is larger for female students compared to male students, and for grade 4 students compared to those in grades 5 and 6; in addition, urban students exhibit larger effects compared to their rural peers, as do students with high use of the platform and high math achievement in the previous year and those who connected during the baseline test. These groups also display a high level of use in the control group. Importantly, in all these subsamples there is a consistent pattern of larger estimated effects on the extensive margin for the Personalized Reminder treatment compared to both the Streaks and the Generic Reminder treatments. Appendix Tables A.8 and A.9 examine heterogeneous effects on the intensive margin (percent of weeks connected for students that connected at least once). Results

[15] 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), For example, while 6 percent of students in the control group achieved a 6-week streak, the corresponding level for students in the Streak treatment was 17 percent.

indicate similar effects of the Streaks treatment across the various subsamples. Still, we do observe a consistent pattern of larger estimated effects on the intensive margin for the Streaks treatment compared to both the Personalized Reminders and the Generic Reminder treatments. As we discuss in Appendix A.2 on external validity, the consistent evidence on the extensive and intensive margins across subsamples suggests that the main qualitative findings of the 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 attempted more exercises compared to students in the control group, an effect of 7.8 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. However, Panel B of Table 1 indicates that pre-treatment characteristics are balanced across treatment groups and the control group for this subsample of students. It is also important to note that this endline test was closely aligned with

the material presented during treatment and does not represent a broad-based measure of academic achievement.

Table 3 reports the estimated effects of our treatments on the fraction of exercises answered correctly in the endline using several alternative specifications: out of the total number of exercises (columns 1 and 2) or the number of 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. These estimates should be interpreted as intent-to-treat (ITT) effects because they do not account for differential take-up. 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.

It is important to recognize that 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.

To further explore the effects of the treatments on academic achievement, we leverage variation in the topics to which students were exposed. That is, we estimate regressions where the dependent variable is standardized math achievement in a specific topic (numeracy, geometry, probability, or patterns) and include an indicator for whether the student connected in a week when a concept was covered. Table 4 presents estimates associated with specification (3) 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. Since students that connected more frequently to the platform tend to have higher baseline academic achievement, we also control for the intensity of platform connections by adding indicators for the number of weeks that students connected, and we further 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). 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, suggesting that the patterns in Panel A are not due to confounding factors. Together, these findings confirm that the actual use of the learning platform generated improvements in math achievement.[16] 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 sending personalized reminders to students. Moreover, highlighting streaks generated larger effects on the fraction of weeks connected for students that connected at least once (the "intensive margin") compared to personalized reminders. Looking at overall measures of platform use that combine both margins, such as percent of weeks connected or 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 for 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

[16] Appendix Table A.12 presents the results from estimating specification (2) in which each topic is regressed separately on a set of indicators for whether students connected during the week(s) when each topic was covered.

is that many individuals start engaging in a desired activity. But highlighting streaks may be a better strategy to achieve higher engagement later, for those that already exhibit the desired behaviors. These strategies could also be implemented differentially over time for different individuals (i.e., for those who have not yet started with the desired behavior, use personalized reminders; but once they have exhibited the desired behavior once, switch to emphasizing streaks).

Our study also documents that highlighting streaks produced improvements in math achievement on 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 does translate to learning gains.

Overall, we believe that the use of streaks can be a valuable strategy to promote the frequency of a desired behavior. 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 into optimal communication strategies.

References

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." Working Paper IDB-WP-304. Washington, DC, United States: Inter-American Development Bank. https://doi.org/10.18235/0001777.

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 28581. Cambridge, United States: 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.

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.

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

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.

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: 3–17. https://doi.org/10.1016/j.jebo.2014.09.005.

Lavecchia, A., H. Liu, and P. Oreopoulos. 2016. "Behavioral Economics of Education: Progress and Possibilities." In: *Handbook of the Economics of Education*. Amsterdam, The Netherlands: 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, United States: National Bureau of Economic Research. https://doi.org/10.3386/w22107.

Mansur, O. 2022. "The Habit-building Research behind Your Duolingo Streak." Duolingo Blog. https://blog.duolingo.com/how-duolingo-streak-builds-habit/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.

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.

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

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 Students 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:

		Connected the previous week	
		Yes	No
Connected this week	Yes	Great job Ana! 🎉 You've accumulated 3 consecutive weeks of practicing mathematics. Let's continue this Monday for more!	Welcome, Ana! 🎊 You've started accumulating weeks in Conecta Ideas. Let's continue this Monday for more!
	No	Hello Ana! 👋🏻 Today, you could reach 3 consecutive weeks of practicing mathematics. We're waiting for you!	Hello Ana! 👋🏻 Today, you could start accumulating weeks of practice in Conecta Ideas. We're waiting for you!

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.[17]

We begin with Appendix Table A.13, 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 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 five cases the difference is statistically significant at the 10 percent level). Finally, we document that in none of the 13 subsamples the

[17] 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 percent of 15-year-olds in Peru reported having a smartphone at home, a figure that rose to 81 percent by 2018. In comparison, the average for developing countries participating in PISA was higher, starting at 89 percent in 2015 and reaching 95 percent by 2018.

overall effects on percent of weeks connected are statistically significantly different between the Personalized Reminder treatment and the Streaks treatment.[18]

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.14 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.

[18] Results available upon request.

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

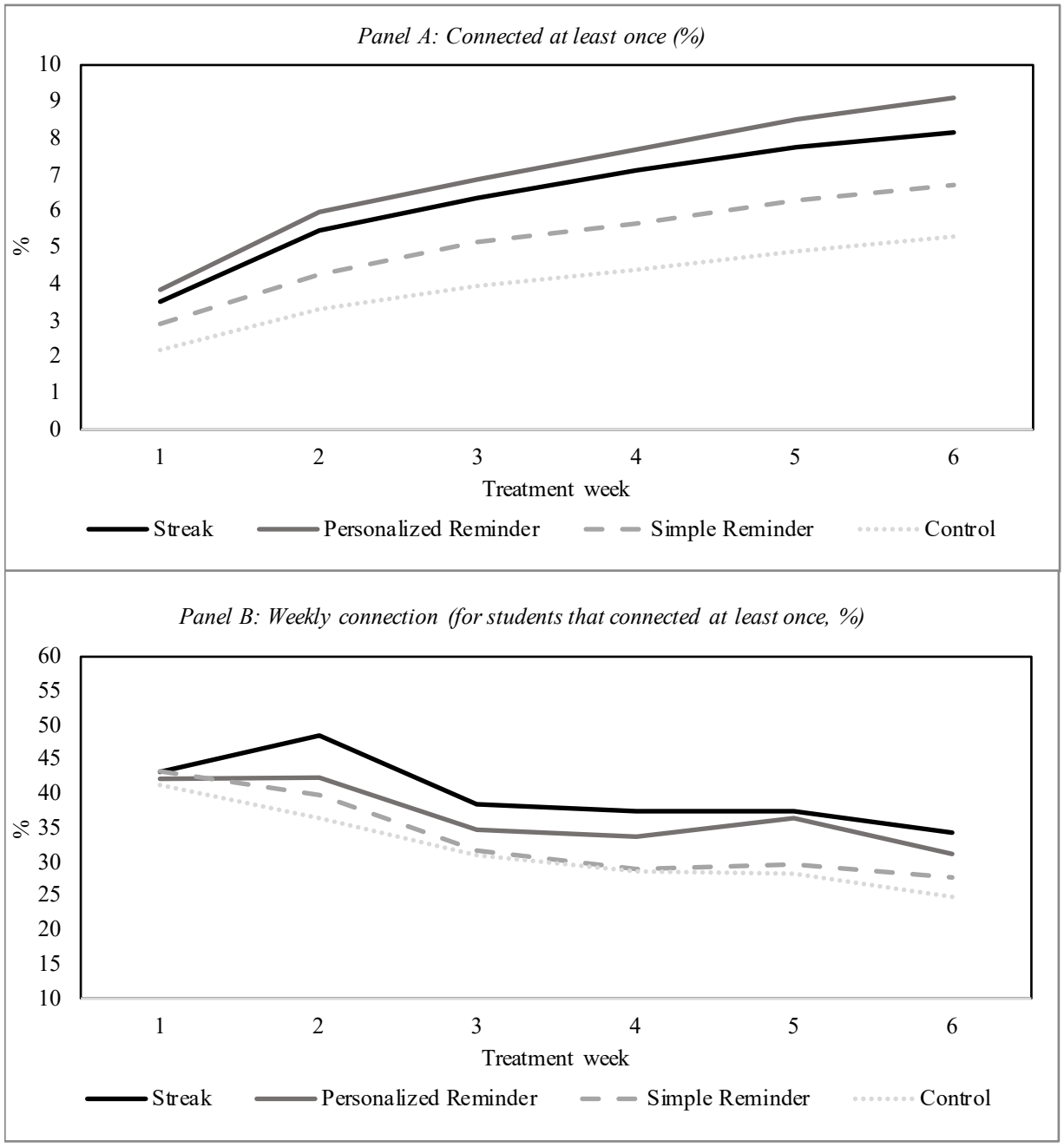
Notes: This table presents estimated effects of the individual and pooled treatments on math achievement. Each column in a panel corresponds to a separate OLS regression. Labels in rows correspond to independent variables. Math achievement in columns (1) and (2) is computed considering all 30 questions 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 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 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 number of weeks connected fixed-effects, specific week fixed-effects, and student fixed-effects, respectively. 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 the endline 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 the endline 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: Association of Exposure to a Topic and Math Achievement

	Topic covered			
	Numeracy	Geometry	Probability	Patterns
	(1)	(2)	(3)	(4)
Panel A: Association with endline test				
Connected for Numeracy	0.42***	0.10	0.29**	-0.04
	(0.13)	(0.13)	(0.13)	(0.13)
Connected for Geometry	0.07	0.22***	0.03	0.12
	(0.08)	(0.08)	(0.08)	(0.08)
Connected for Probability	0.03	0.22**	0.17*	0.05
	(0.09)	(0.09)	(0.09)	(0.09)
Connected for Patterns	0.10	-0.03	0.08	0.19**
	(0.09)	(0.09)	(0.09)	(0.09)
N	*935*	*935*	*935*	*935*
Panel B: Association with baseline test				
Connected for Numeracy	0.20*	0.14	0.20*	0.17
	(0.11)	(0.11)	(0.11)	(0.11)
Connected for Geometry	0.19***	0.10	0.17**	0.06
	(0.07)	(0.07)	(0.07)	(0.07)
Connected for Probability	0.16*	0.16*	0.09	0.10
	(0.08)	(0.08)	(0.08)	(0.08)
Connected for Patterns	0.01	0.05	0.02	-0.01
	(0.08)	(0.08)	(0.08)	(0.08)
N	*1,156*	*1,156*	*1,156*	*1,156*

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 a student. Each column in a panel corresponds to a separate OLS regression. The dependent variable in Panel A is math achievement in a specific subject in the endline test. The dependent variable in Panel B is math achievement in a specific subject in the baseline test. Column (1) presents coefficients and standard errors of a regression of math achievement in numeracy on exposure to numeracy, geometry, probability and patterns during the six-week treatment period. Exposure to numeracy is computed as the share of weeks that the student connected when this topic was covered. Exposure to geometry is represented with a dummy variable that equals one if the student connected to the platform when this topic was covered (exposure to probability and patterns are constructed in a similar fashion). Columns (2) to (4) presents the results of similar regressions but with dependent variables corresponding to math achievement in geometry, probability and patterns, respectively. Dummies for number of weeks connected are included in all regressions. 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.

Table A.13: 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.14: 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.