Discovering Data-Driven Nudges to Help Students Learn More Math

Marcos Gallo Sean Hu Ben Manning Angela Duckworth
Colin Camerer

Introduction

The decline in math performance among American students has been a critical issue exacerbated by the COVID-19 pandemic, with some reporting an alarming half-year lag in math achievement among U.S. public school students in grades 3-8 (Fahle et al. 2023). Learning losses and disparities, particularly among younger students, have been significant due to reduced instructional time and remote learning challenges (Zierer 2021; Di Pietro 2023). Moreover, Ewing and Green (2021) note that repeated school closures compounded these issues, emphasizing that the pandemic is not the only problem affecting math performance. Historical trends have also played a role, indicating a need for comprehensive solutions.

Emphasizing quality math instruction is paramount for improving math performance. Teacher proficiency in pedagogical strategies, guidance, and teacher-student relations is crucial, especially for students who have lost interest in mathematics or lack a sufficient foundation (Wang et al. 2023; Battey et al. 2016). The correlation between teachers' deep mathematical knowledge and student success highlights the importance of specialized training (Hill et al. 2008; Battey et al. 2016).

The pandemic also accelerated the adoption of digital platforms for math education. Indeed, adaptive practice software can mitigate the impact of school closures and possibly reverse their adverse effects (Meeter 2021; Alabdulaziz 2021). Recent meta-analyses demonstrate that integrating digital tools and blended learning approaches improves student outcomes significantly (Ran, Kim, and Secada 2021; Ran, Kasli, and Secada 2020; Sadaf, Wu, and Martin 2021). Leveraging these tools and insights gained during the pandemic should help address longstanding educational challenges (Ewing and Green 2021). In particular, integrating technology, pedagogy, and content knowledge is essential, and professional development programs are most effective when they focus on using technology to foster a more engaging and effective learning environment (Young 2016; Blanchard et al. 2016).

The rise of digital platforms has also underlined the importance of student engagement in online learning. Blending online and traditional teaching methods effectively enhances engagement and understanding (Chiang, Lin, and Tseng 2016; Sadaf, Wu, and Martin 2021). Teachers are essential in helping students develop meta-cognitive skills to enhance student engagement (Haleva, Hershkovitz, and Tabach 2020). Furthermore, teachers' beliefs and self-efficacy toward technology integration influence their willingness to adopt innovative teaching practices (Ertmer et al. 2012; Liljedahl and Oesterle 2020). This relationship between student engagement and teacher attitudes highlights the importance of providing cost-effective, high-quality solutions that can be implemented across diverse educational settings.

In this study, we partner with Zearn to address these topics. This math education platform reaches approximately 25% of United States elementary schools and over a million middle-school students. Zearn's approach integrates interactive digital lessons with hands-on teaching, aligning with the Common Core State Standards and providing a comprehensive educational experience ("Zearn Math: Top-Rated Math Learning Platform" 2023).

Our study leverages this rich resource to offer an innovative approach to educational interventions using behavioral science principles. Focusing on teacher quality, we align with Hanushek (2020), who maintains that the efficacy of resource utilization supersedes quantity. We also follow current trends in providing cost-effective, easy-to-implement interventions that align engagement incentives (Lavecchia, Liu, and Oreopoulos 2016; Koch, Nafziger, and Nielsen 2015).

Our two-step approach initially employs unsupervised learning techniques to analyze behavioral patterns in teacher activity on Zearn, aligning with the data mining value in educational research (Salazar, Serrano, and Vergara 2007; Hershcovits, Vilenchik, and Gal 2020; Qiu et al. 2022; Al-Shabandar et al. 2018; Shin and Shim 2020). Subsequently, we aim to establish the causality of our interventions through a large-scale field experiment guided by recommendations from Greene (2022) for holistic, transparent, and reproducible research. Our unique integration of behavioral science with digital education seeks to provide impactful insights into math education and offer a blueprint for similar studies in other fields.

Results

Study 1: Data-Driven Nudge Engineering

Our study used a comprehensive dataset from Zearn's educational platform, encompassing the 2019-2020 academic year and spanning multiple schools in Louisiana. Zearn's approach combines concrete, pictorial, and abstract methods for teaching mathematics, offering a personalized experience that allows teachers to track student progress effectively. Key features like the Badge system, which tracks student lesson completion, and Tower Alerts, which notifies teachers when students repeatedly struggle with a given concept, motivate students and provide valuable insights for educators.

Key variables from the dataset included teacher logins, file downloads, and specific interactions with educational content. Additionally, we accessed alongside student data at the classroom-week level, encompassing metrics such as lesson completion (i.e., Badges) and instances of learning difficulties (i.e., Tower Alerts). This granularity allowed for an in-depth analysis of both teacher behaviors and student performance.

The data revealed diverse engagement patterns. Teachers logged into Zearn multiple times weekly, with notable variations in interaction frequency and duration. Student data reflected a broad spectrum of performance levels across different classrooms and schools. The standard deviations of key variables underscored this diversity, as seen in Table 1, which displays comprehensive summary statistics. This rich combination of teacher behaviors and student performance metrics, carefully matched with a weekly frequency for each classroom, allowed for a thorough analysis while upholding privacy standards.

Table 1: Pre-experimental Intervention Data. This table summarizes key educational metrics for Zearn teachers. The data was collected from July 2019 to June 2020.

		Aggregated Statistics per Teacher				
	N	Mean	Standard Deviation	1st Quartile	Median	3rd Quartile
# of teachers	2,288	-	-	-	-	-
# of classes	3,807	1.66	0.94	1.0	1.0	2.0
# of badges	5,118,450	2,237.08	1,610.13	1,198.8	1,840.7	2,788.4
Total minutes on Zearn	4,280,790	175.00	717.76	3.8	19.9	97.4

Study 1a

Zearn offers a variety of student and teacher activities, which, in our data, can be simplified through dimensionality reduction techniques, enhancing interpretability. We employed Independent Component Analysis (ICA) on teacher behavioral variables due to their non-Gaussian nature. This statistical approach allowed us to uncover latent variables that may have been obscured with traditional methods. These components are weighted vectors of specific teacher activities and were estimated to maximize statistical independence in the data-generating process. Our findings revealed three pivotal independent components, as indicated by the 'elbow' of Figure 1, which represents the optimal number of components that best portray the underlying dimensions of teacher behavior. (For more details, please refer to the SI section.)

The significant finding from the ICA was the prominence of IC1, accounting for 14.56% of the variance in teacher Zearn activity, as indicated in Figure 2. This result is significant in nudge engineering, highlighting the need to focus interventions on elements encapsulated by IC1. Conversely, IC2 and IC3, with 12.61% and 6.16% variance explained, play lesser but still noteworthy roles.

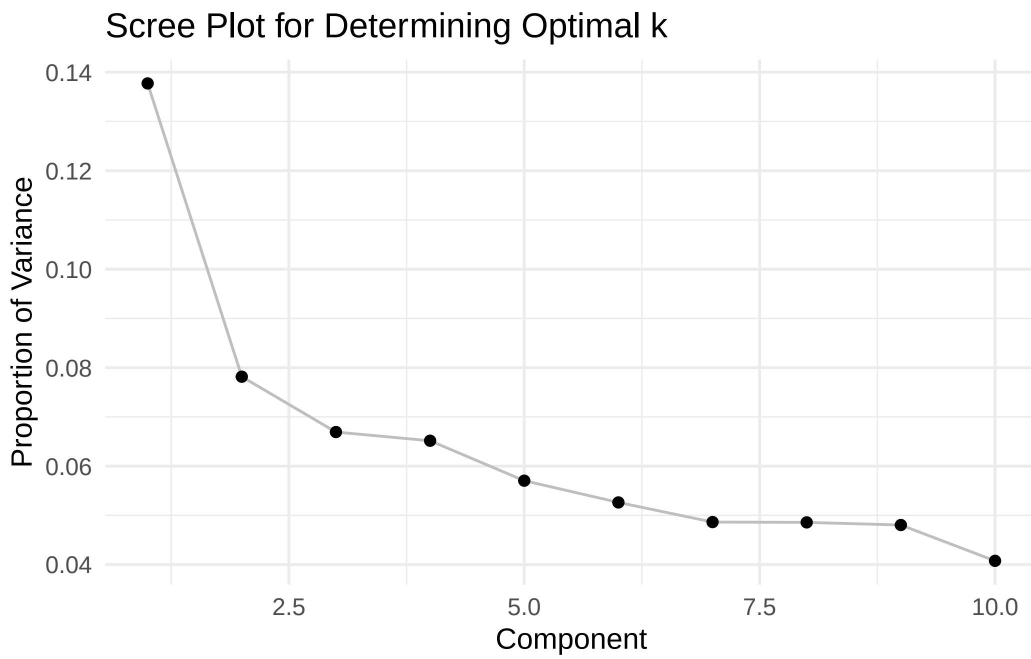

Figure 1: Scree Plot for Determining Optimal Number of Components. This plot displays the proportion of variance explained by each independent component (IC). The optimal number of components is indicated by the 'elbow' of the plot, where the variance explained by each additional component is minimal.

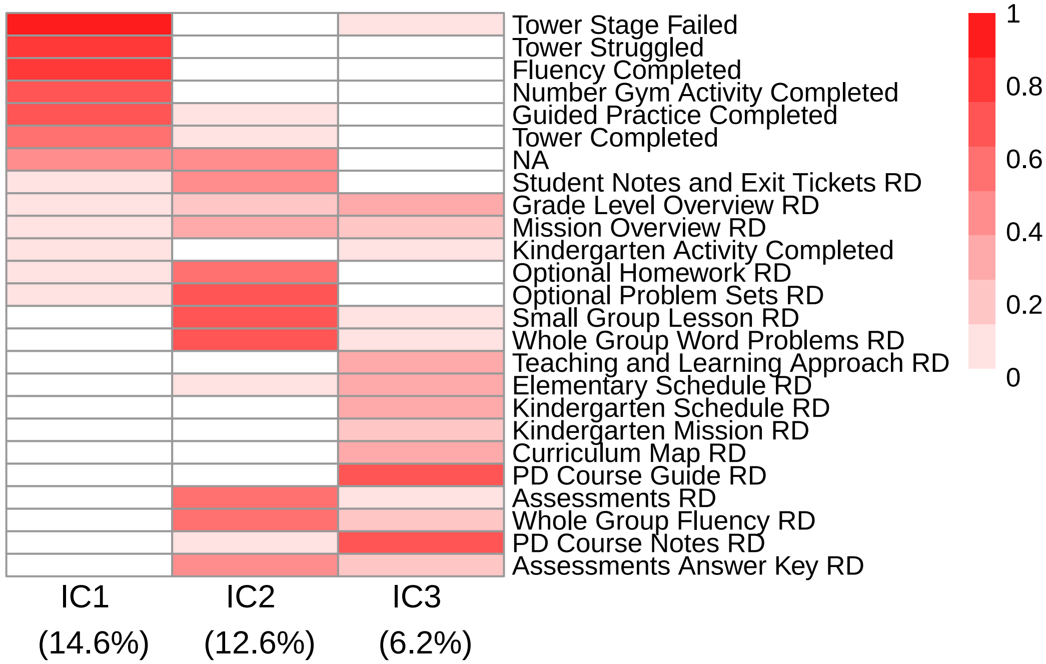

Figure 2: Independent Component Analysis (ICA) Results. The heatmap displays teacher actions in each row, while the columns represent the three independent components (IC1, IC2, IC3) that explained the most variance in the teacher behavioral data, with their respective percentage of variance explained in parentheses. The color gradient on the heatmap indicates the relative importance of each activity within these components. Note that these metrics pertain to teacher activity on the platform, not student actions.

The activities with the highest weighting in IC1 were measures associated with Towers (Struggled, Failed, and Completed) in conjunction with other completed student activities such as Fluency, Guided Practice, and Number Gym Activity. Notably, these metrics pertain to teacher logins, not student actions, implying a direct link between teacher engagement and student performance, suggesting an educational "scaffolding" effect.

Further examination revealed an empathy-driven process behind this trend. In this context, "empathy" refers to teachers' ability to comprehend and engage with students' challenges, as demonstrated by their focused attention on areas where students struggled or succeeded. The weightings in IC1 for activities like Tower Struggles (0.871), Fluency Completed (0.846), Guided Practice Completed (0.652), and Number Gym Activity Completed (0.655) were significantly higher compared to other activities. This pattern underlines the empathy-driven process in teaching, where teachers' engagement with the platform is focused on understanding student challenges.

The second component, IC2, showed substantial weightings on variety of Resource Downloads (RD), particularly Small Group Lessons (0.719) and Whole Group (i.e., the entire class) Word Problems (0.65). We labeled this component as "Groups." IC3, with strong weights on Resource Download (RD) for the Professional Development (PD) course guide (0.643) and course notes (0.638), was labeled "Guidance." This component, accounting for a smaller variance (6.159%), was harder to interpret due to its complexity and diverse activities.

Building upon these findings, we advanced to a censored (lower bound = 0) panel regression model that accounted for various temporal and subject-specific variables and the potential impact of teachers handling multiple classes. This model applied a first-difference approach with a lower bound of zero for the dependent variable. The regression formula incorporated changes in independent components (IC1, IC2, IC3) as predictors for the change in the logarithm of badges (+1), accounting for individual teachers, classes, and weeks. As presented in Table 2, the regression highlighted a strong positive contemporaneous correlation between IC1 and badges, with a coefficient of 0.0266 ($p < 0$), suggesting a significant impact on student achievement.

Table 2: Regression of Log Badges on the Independent Components. This table displays the results of a censored panel regression model with a lower bound of 0 for the dependent variable. Standard errors are clustered at the teacher level.

Variable	Coefficient	Std. Error	t value	P Value
teacher_number_classes	−0.0204	0.0026	−7.8598	0.0000
Grade.Level.L	0.3839	0.0564	6.8055	0.0000
Grade.Level.Q	−0.0520	0.0292	−1.7793	0.0752
Grade.Level.C	0.2854	0.0594	4.8075	0.0000
Grade.Level^4	0.1024	0.0632	1.6197	0.1053
Grade.Level^5	0.1941	0.0639	3.0396	0.0024
Grade.Level^6	0.0442	0.0774	0.5704	0.5684

Grade.Level^7	0.0723	0.0664	1.0892	0.2761
Grade.Level^8	−0.0389	0.0348	−1.1185	0.2633
Students...Total	0.0011	0.0008	1.3226	0.1860
ic1	0.0266	0.0129	2.0650	0.0389
ic2	0.0818	0.0041	20.0092	0.0000
ic3	−0.0093	0.0019	−4.8522	0.0000

The correlation we discovered has significant implications. It indicates that an increase in activities associated with IC1, such as encountering one more 'Tower Struggle' than average, can lead to an approximate 4.17 percent increase in student badges. Although this increase may seem small, it may be substantial in the context of complex human behaviors, where even modest changes often lead to far-reaching effects. This observation highlights the potential of targeted interventions based on our findings to yield significant improvements in educational outcomes.

Study 1b

Prompted by Buyalskaya et al. (2023), we sought to uncover the subtleties of habitual behaviors within an educational setting. Our primary goal was to understand how regular teacher interactions with the Zearn platform impacted student learning outcomes. To measure this, we used the log-transformed average weekly badges per student over the entire school year as our dependent variable. We constructed our explanatory variables with careful consideration of the patterns that could identify habitual engagement and their relationship to student performance:

1. Login Percentages across months and days of the week: The frequency of logins across different time periods. For example, among all the logins for a teacher, we assess how many are from January or Tuesdays, compared to all other months and days of the week, respectively. Note that our regression omitted July and Sunday, periods with low login incidence.
2. Time Spent on Zearn: Teachers' weekly average time spent on the platform, measured in minutes.
3. Average Streak: Average number of consecutive weekdays in which the teacher logs in.
4. Average Days Between Logins: Average number of days between two login instances.

To account for any school-specific factors that may have influenced the relationship between teacher behavior and student achievement, we estimated a school-fixed-effects regression model. Unlike study 1, this regression did not follow each teacher or class across weeks, as our unit of analysis was a teacher summed across classrooms and averaged across all weeks.

The regression results, as detailed in Table 3 (see Table S2 for full results), revealed that all weekday login percentages positively affected student badges. However, Friday logins stood out significantly, supporting our hypothesis that specific days of the week have more influence on habitual engagement.

Table 3: Regression of Log Average Badges on Habit Variables. This table displays the results of the regression model with school fixed effects. Coefficients on months and days measure the difference between the effects of login percentage relative to the July percentage (for months) and the Sunday percentage (for days). Standard errors are clustered at the school level.

Variable	Coefficient	Std. Error	t value	P Value
%Monday	0.0210	0.0102	2.0523	0.0402
%Tuesday	0.0140	0.0073	1.9316	0.0535
%Wednesday	−0.0076	0.0063	−1.1931	0.2329
%Thursday	0.0182	0.0062	2.9275	0.0034
%Friday	0.0290	0.0075	3.8375	0.0001
%Saturday	0.0131	0.0125	1.0462	0.2955
%January	−0.0092	0.0084	−1.0858	0.2776
%February	−0.0104	0.0067	−1.5578	0.1194
%March	−0.0011	0.0062	−0.1739	0.8619
%April	−0.0202	0.0063	−3.2049	0.0014
%May	−0.0001	0.0080	−0.0085	0.9932
%September	−0.0192	0.0062	−3.0919	0.0020
%October	−0.0159	0.0082	−1.9392	0.0525
%November	−0.0148	0.0123	−1.1987	0.2307
Avg. Minutes	0.0059	0.0029	2.0500	0.0404
Avg. Streak	−0.0549	0.1711	−0.3205	0.7486
Avg. Days Between Logins	0.0299	0.0194	1.5371	0.1244

In particular, our analysis indicates that shifting 10% of logins from other weekdays to Fridays, without increasing overall platform usage, could boost student lesson completion by around 17.56%. For the typical teacher, this means switching from one Friday login per month to two while reducing one login from another weekday during that month, resulting in an increase of 26.86% in average weekly lesson completion. This outcome underscores the importance of strategic engagement rather than just login frequency.

We attribute this pronounced effect to two key factors. First, Friday logins facilitate "Reflective Catch-Up," enabling teachers to review and analyze the previous week's activities and make necessary adjustments. Second, "Foresighted Planning" occurs on Fridays as teachers proactively plan for the upcoming week, a practice less common on weekends.

Study 2: Nudge Engineering - From Data to Intervention Design

In Study 1, we discovered that specific psychopedagogical strategies, habits, and timing of teacher interactions were crucial in promoting student success on digital platforms. With this foundation, we could craft effective educational interventions and strategies. By leveraging the insights from the first study, we aimed to develop targeted interventions that maximize student learning outcomes.

Study 2a

We used the first component from the ICA in Study 1a to design an "empathy" nudge. This intervention involved sending emails to teachers, encouraging them to adopt a more student-centered teaching approach by viewing math problems from their students' perspective. Our analysis in Study 1a indicated that this process was linked to the highest weighted behaviors in IC1. The emails contained key messages emphasizing the importance of empathy in teaching math, advice from other teachers, and helpful tips for assisting students who struggle with a lesson. The emails also suggested specific actions, including using Zearn's features to view lessons from a student's perspective and checking the Tower Alerts Report (see SI for the complete emails). We hypothesized that this empathy approach would significantly enhance student performance.

Study 2b

In Study 2b, we aimed to test whether nudging teachers to log in on Fridays, as opposed to an unspecified day, would improve student performance. Our approach involved sending emails to teachers, encouraging them to log in on Fridays, highlighting effective teaching habits, and emphasizing the benefits of regular check-ins on student progress. These emails included testimonials from other teachers, research insights, and motivational messages to encourage habit formation (see SI for the complete emails). Our rationale was that Friday logins would aid in reflecting on the week's activities and proactive planning for the following week.

In addition, we designed a control email tailored to this treatment. These emails were sent every Wednesday to remind Zearn teachers to review their Zearn Pace Report without the personalized behavioral prompts or motivational materials in the Friday treatment group emails. Although the control emails prompted participation with Zearn, they did not provide information on Friday logins, introspection, or strategic preparation.

We hypothesized that the Friday approach to teacher engagement with the platform would yield a greater impact on student achievement than the Wednesday control or our study-wide control.

Impact Evaluation

In collaboration with Zearn, Study 2 became part of a large multi-arm "megastudy," set in motion during a critical four-week period in 2021, involving over 140,000 teachers and nearly 3 million students. Teachers in our study taught a median of 20 students (mean = 21.30, SD = 15.31) in a median of 1 classroom (mean = 1.15, SD = 0.59). Before the intervention, teachers in our study had, on average, logged into Zearn a total of 3.62 times between July 1, 2021, and September 14, 2021 (SD = 8.49) (see SI for a complete description of the sample).

Table 4: Efficacy of Different Teacher Engagement Interventions on Student Learning Outcomes. The table showcases the effects of the 'Empathy' and 'Friday Login' interventions compared to the study-wide control, along with the results from the Friday-specific control group. We measure student achievement by the number of badges students earned.

Treatment	Coefficient	Std. Error	t value	P Value
Empathy	0.0721	0.0304	2.3748	0.0176
Friday	0.0550	0.0301	1.8250	0.0680
Friday-control	0.0898	0.0299	2.9988	0.0027
Intercept	2.0056	0.0343	58.5501	0.0000

As stated in our preregistration, we evaluated the impact of intervention messages on the number of math lessons completed by students during the four-week intervention period. Students in the megastudy control condition completed a regression-estimated 1.761 lessons during the 4-week intervention period. Table 4 shows that our interventions increased the number of math lessons completed by students during the intervention period by a regression-estimated average of 0.0487 lessons, which is a 2.77% increase over the megastudy control condition. Specifically, the empathy treatment increased the number of lessons completed by students by 0.0721 lessons, or a 4.09% increase over the megastudy control condition (d = 0.0188, p = .018). The Friday treatment increased the number of lessons completed by students by 0.0550 lessons, or a 1.81% increase over the megastudy control condition (d = 0.0192, p = .068). The Friday treatment was not significantly different from the Friday-specific control (F(1,118137) = 0.85, p = .357).

Table S3 presents unstandardized coefficients from our primary regression analysis and unadjusted, robust SEs and CIs. Additionally, we utilized the Benjamini-Hochberg (BH) procedure to compute adjusted p-values, which help to control for false discovery rates when conducting multiple comparisons (Benjamini and Hochberg 1995). Before adjusting for multiple hypothesis testing, all treatments exhibited significant benefits. However, only our treatment-specific control had a BH-adjusted p-value of less than 0.05. This intervention involved encouraging teachers to log in to Zearn weekly to receive updated student performance reports. Although reliable, the effect of this intervention was small, resulting in an estimated 0.0898 extra lessons completed in four weeks, or a 5.10% increase over the megastudy control ($d = 0.0235$, $p = .003$). Even after applying the James-Stein shrinkage procedure to adjust for the winner's curse (i.e., the maximum of 15 estimated effects is upward biased), we estimated that this intervention still produced 0.061 extra lessons completed, or a 3.46% increase over the control condition (James and Stein 1992).

Discussion

This study aimed to improve student math learning on the Zearn platform by integrating data analysis into educational intervention. We identified critical teacher behaviors influencing student performance and evaluated two novel interventions: empathy and Friday habitization.

Our first study revealed subtle but significant patterns in teacher engagement that traditional analyses might overlook. Study 1a identified a significant independent component strongly associated with metrics related to struggles and achievements, suggesting teachers' empathy-driven engagement, focusing on areas where students faced challenges or succeeded. This pattern also correlated with a significant increase in student achievement, aligning with findings emphasizing the importance of teacher empathy in educational outcomes (Hill et al. 2008; Battey et al. 2016). In Study 1b, we discovered that teachers who logged into Zearn on Fridays had a notable impact on student math performance. This behavior indicated a commitment to continuous planning and support, echoing the findings of Blanchard et al. (2016) that technology integration does not need large-scale changes in practices to enhance student learning. Overall, our research highlights the importance of teacher engagement and personalized instruction and feedback in improving student performance on the Zearn platform. As Ertmer et al. (2012) have emphasized, aligning student-centered beliefs and practices is vital to success, regardless of technological, administrative, or assessment barriers.

In Study 2a, the success of the empathy intervention can be attributed to its alignment with psychological principles that emphasize the importance of emotional connectivity between teachers and students. This intervention likely fostered a more engaging and supportive learning environment, a feature essential for the success of digital platforms (Lavecchia, Liu, and Oreopoulos 2016; Koch, Nafziger, and Nielsen 2015). In contrast, Study 2b's unexpected results highlight the complexities of behavior change in educational settings, suggesting that repetitive routines without meaningful engagement or context may not enhance learning outcomes. Notably, our study-specific control outperformed all other megastudy interventions. Initial analyses from Duckworth et al. (2024) suggest that this effect is due to the higher salience of personalization present, such as the suggestion of classroom-specific actions (e.g., "CLICK HERE to see which of your students are struggling"). The lack of additional content in this control highlighted actionable steps to engage with students, an effect which, in retrospect, aligns with previous literature.

Our study's insights transcend the immediate context. It showcases the potential of data-driven interventions to create strategies that cater to the unique needs of teachers and students. This approach paves the way for personalized and responsive pedagogy. In the realm of digital learning, our findings underscore the pivotal role of teacher engagement and tailored content, which are crucial for replicating and enhancing the benefits of traditional classrooms. In essence, our research provides a roadmap for designing online educational tools that are more effective and engaging.

Our study, while insightful, is limited by its focus on a specific demographic and educational context within the Zearn platform. It also only partially captures teacher-student interactions in traditional classroom settings. Consequently, generalizing our results to other educational settings, cultures, or age groups may be challenging. Additionally, while our approach was more cost-effective and less time-intensive, it achieved a more modest impact than the substantial effects seen in more intensive programs (Banerjee et al. 2007; Di Pietro 2023). The simplicity of our email interventions and the short duration of the study likely contributed to these results, although their magnitude aligns with other reports from educational technology applications (Cheung and Slavin 2013). Future research should explore more engaging and intensive intervention methods over extended periods to potentially yield greater impacts on learning outcomes.

While rooted in education data, our research introduces a paradigm with significant implications beyond its primary focus. By combining data-driven analysis with targeted behavioral interventions, our approach offers a flexible framework that can be adapted to various fields. Whether in healthcare, environmental behavior, or organizational management, our methodology demonstrates the potential to harness data insights for effective behavioral change. Hence, our study also serves as a catalyst for innovative approaches in diverse fields where behavior modification is crucial.

Materials and Methods

Study 1

Data Collection

The Zearn math educational learning platform provided us with administrative data spanning the 2019-2020 academic year (September 2019 to May 2020). This dataset included detailed teacher actions and classroom-week-level student achievement metrics. Teacher actions on the platform were timestamped to the second. For privacy considerations, Zearn aggregated student data at the classroom-week level, including student achievement measures and indications of student struggles. The dataset covered various schools across Louisiana (**?@fig-teachers-map**).

To promote transparency and replicability of our study, we have deposited the data and code used in our analyses in a publicly accessible database. Researchers and interested parties can access the complete dataset and all related processing scripts at the GitHub repository: https://github.com/SeanHu0727/zearn_nudge.git

Inclusion Criteria

We aggregated teacher behavior data to the weekly level and merged it with the student data at the classroom-week level. We also excluded inactive teachers (those with no recorded activity for over two months) from the dataset.

We defined the inclusion criteria for classrooms strictly as such:

1. Classrooms linked to a single teacher
2. Classrooms with no more than seven months of inactivity during the academic year
3. Classrooms with an average of no less than five actively engaged students

Study 1a

Independent Component Analysis (ICA)

We extracted all teacher behavioral variables from the dataset that displayed non-zero variance and standardized them to have a zero mean and unit variance. To determine the ideal number of independent components, we performed ICA using a range of components from 1 to 10. Our decision on the optimal number was informed by recognizing the 'elbow' on the scree plot generated from the results, yielding three independent components. We used the `icafast` function from the R `ica` package for all ICAs conducted (Helwig 2022).

Censored Panel Regression

We estimated a censored panel regression model using a first difference approach, with the `pldv` function from the `plm` package (Croissant and Millo 2008) in R and a lower bound of 0 for the dependent variable:

$$\Delta \ln(\text{Badges}_{itc} + 1) = \beta_0 + \beta_1 \Delta(\text{IC1})_{it} + \beta_2 \Delta(\text{IC2})_{it} + \beta_3 \Delta(\text{IC3})_{it} + \Delta\epsilon_{itc}$$

where i, c, t index the teacher, class, and week, respectively. Standard errors were clustered at the teacher level with the `vcovHC` function (Millo 2017) in R.

Study 1b

Fixed-Effects Regression

We estimated a fixed effects model using the `plm` function(Croissant and Millo 2008) in R:

$$\begin{aligned}\ln\left(\sum_{c=1}^{C_i} (\text{Avg. Badges})_c + 1\right) = \beta_0 &+ \sum_{m=1, m\neq 7}^{12} \beta_m \text{Login\%}_{m,i} + \sum_{d=1, d\neq 7}^{7} \beta_{12+d} \text{Login\%}_{d,i} \\ &+ \beta_{19}(\text{Avg. Minutes})_i + \beta_{20}(\text{Avg. Streak})_i \\ &+ \beta_{21}(\text{Avg. Days Between Logins})_i + \text{School FEs} + \epsilon_i\end{aligned}$$

for teacher i with C_i classes. `Login%` represents the percentage of logins by teacher i during each month m and on each day of the week d relative to other months and days, respectively. We exclude July and Sunday to prevent multicollinearity. Standard errors are clustered by school with the `vcovHC` function (Millo 2017) in R.

Study 2

We used the findings from Study 1 to inform the creation of two interventions as part of a larger multi-arm "mega study" that involved 15 sets of nudges.

Implementation

Study 2 was conducted in collaboration with Zearn ("Zearn Math: Top-Rated Math Learning Platform" 2023) and was preregistered for the fall of 2021. To incentivize teacher participation during our "intervention period" from September 15, 2021, to October 12, 2021, all teachers on the platform received two messages on September 1 and 8, 2021. These messages informed them they had been enrolled in the "Zearn Math Giveaway" and that every email opened until October 12, 2021, would earn them tickets. These tickets were used to enter drawings for various prizes, such as autographed children's books, stickers, and gift cards.

Data

We excluded Zearn elementary school teachers from the study if they (a) taught grades other than first through eighth ($n = X$), (b) lacked a valid email address in the Zearn system as of September 8, 2021 ($n = X$), (c) had less than one or more than 150 students associated with their Zearn account as of October 18, 2021 ($n = X$), (d) had fewer than one or more than six classrooms associated with their Zearn account as of DATE ($n = X$), (e) had other teachers associated with their Zearn classroom(s) as of DATE ($n = X$), or (f) had not logged onto the Zearn platform (or had no associated student who logged onto the platform) between March 1, 2021 and September 14, 2021 ($n = X$).

After exclusions, we randomized $N = 140,461$ teachers across 22,281 schools who served 2,992,077 students in 161,722 classrooms into one of the intervention conditions or the control condition ($N_{\text{control}} = 29,513$). The control condition was larger than the interventions to account for multiple comparisons. Among $n = 16,372$, or 11.66%, of teachers, at least one of two problems occurred in the emails sent by Zearn Math during the intervention period: an email message that was intended but not sent ($n = 13,568$, or 9.66% of teachers), or an email message that was sent but not intended (i.e., from a different treatment condition; $n = 2,804$, or 2.00% of teachers). As these email problems were systematically related to treatment assignment ($\chi^2 = 33.01$, $df = 15$, $p = .005$), we did not exclude these participants and conducted intent-to-treat analyses. Refer to the SI for email problem prevalence by condition and study analyses that exclude or adjust for email problems, respectively.

Impact Assessment

We followed our preregistered analysis plan to assess the effect of each treatment on the primary outcome of interest: math lessons completed by students during our four-week intervention period (Gallo et al. 2022b, 2022a). We estimated a weighted ordinary least squares (OLS) regression with the `areg` command in Stata (StataCorp 2023). Each teacher's observations were weighted proportionally to the total number of students in their Zearn classroom(s).

The primary predictors were indicators for each intervention, omitting the control condition. The regression also included the following control variables: (1) school fixed effects, (2) an indicator for the teacher's account type (free or paid), (3) the number of times the teacher logged into Zearn prior to the study, from August 1 to September 14, 2021, (4) the total number of students in the teacher's classroom(s) as of October 18, 2021, (5) the number of classrooms associated with the teacher as of October 18, 2021, (6) the number of days since the teacher obtained a Zearn account prior to the study's launch, (7) the number of days separating the study's launch and the start of the teacher's school year, (8) the average number of lessons completed by a teacher's students from the start of their school year to the start of the intervention (or from July 14, 2021, if the school year start was not known), (9) whether the teacher opened our September 1, 2021 email announcing the upcoming Zearn Math Giveaway, (10) a similar indicator for our September 8, 2021, email reminding them of the giveaway, and (11) the percentage of a teacher's students in each grade except for third grade to avoid multicollinearity, since for most teachers, students were in a single grade.

Study 2a: Empathy-Based Intervention Design and Implementation

We designed four emails highlighting empathy and encouraging teacher engagement, specifically regarding pedagogical content knowledge. A group of 7,443 teachers were chosen at random to receive these emails.

Study 2b: Habitization-Based Intervention Design and Implementation

We designed four emails to be sent on Fridays, emphasizing the importance of regular Friday logins and student progress tracking. A total of 7,476 teachers were randomly selected to receive these emails. Additionally, we created a condition-specific control group that received messages on Wednesdays without Friday-specific content but with links to specific actions on Zearn. A group of 7,577 teachers were randomly chosen to participate in this control group.

Ethical Considerations and IRB Approval

This study was conducted in accordance with ethical standards and received exempt status from the Institutional Review Board (IRB) at the University of Pennsylvania. The study's methodologies were designed to ensure the confidentiality and anonymity of all participants involved, adhering strictly to ethical guidelines for educational research. Data received from Zearn was aggregated at the classroom-level, with no identifying information about teachers or students.

References

Alabdulaziz, Mansour Saleh. 2021. "COVID-19 and the Use of Digital Technology in Mathematics Education." *Education and Information Technologies* 26 (6): 7609–33. https://doi.org/10.1007/s10639-021-10602-3.

Al-Shabandar, Raghad, Abir Jaafar Hussain, Panos Liatsis, and Robert Keight. 2018. "Analyzing Learners Behavior in MOOCs: An Examination of Performance and Motivation Using a Data-Driven Approach." *IEEE Access* 6: 73669–85. https://doi.org/10.1109/access.2018.2876755.

Banerjee, A. V., S. Cole, E. Duflo, and L. Linden. 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics* 122 (3): 1235–64. https://doi.org/10.1162/qjec.122.3.1235.

Battey, Dan, Rebecca A. Neal, Luis Leyva, and Karlyn Adams-Wiggins. 2016. "The Interconnectedness of Relational and Content Dimensions of Quality Instruction: Supportive Teacher–student Relationships in Urban Elementary Mathematics Classrooms." *The Journal of Mathematical Behavior* 42 (June): 1–19. https://doi.org/10.1016/j.jmathb.2016.01.001.

Benjamini, Yoav, and Yosef Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1): 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Blanchard, Margaret R., Catherine E. LePrevost, A. Dell Tolin, and Kristie S. Gutierrez. 2016. "Investigating Technology-Enhanced Teacher Professional Development in Rural, High-Poverty Middle Schools." *Educational Researcher* 45 (3): 207–20. https://doi.org/10.3102/0013189x16644602.

Buyalskaya, Anastasia, Hung Ho, Katherine L. Milkman, Xiaomin Li, Angela L. Duckworth, and Colin Camerer. 2023. "What Can Machine Learning Teach Us about Habit Formation? Evidence from Exercise and Hygiene." *Proceedings of the National Academy of Sciences* 120 (17): e2216115120. https://doi.org/10.1073/pnas.2216115120.

Cheung, Alan C. K., and Robert E. Slavin. 2013. "The Effectiveness of Educational Technology Applications for Enhancing Mathematics Achievement in K-12 Classrooms: A Meta-Analysis." *Educational Research Review* 9 (June): 88–113. https://doi.org/10.1016/j.edurev.2013.01.001.

Chiang, Po-Jui, Ya-Wen Lin, and Chih-Lung Tseng. 2016. "The Effect of Blended Learning in Mathematics Course." *EURASIA Journal of Mathematics, Science and Technology Education* 13 (3). https://doi.org/10.12973/eurasia.2017.00641a.

Croissant, Yves, and Giovanni Millo. 2008. "Panel Data Econometrics in *R* : The **Plm** Package." *Journal of Statistical Software* 27 (2). https://doi.org/10.18637/jss.v027.i02.

Di Pietro, Giorgio. 2023. "The Impact of Covid-19 on Student Achievement: Evidence from a Recent Meta-Analysis." *Educational Research Review* 39 (May): 100530. https://doi.org/10.1016/j.edurev.2023.100530.

Ertmer, Peggy A., Anne T. Ottenbreit-Leftwich, Olgun Sadik, Emine Sendurur, and Polat Sendurur. 2012. "Teacher Beliefs and Technology Integration Practices: A Critical Relationship." *Computers & Education* 59 (2): 423–35. https://doi.org/10.1016/j.compedu.2012.02.001.

Ewing, Eve L., and Terrance L. Green. 2021. "Beyond the Headlines: Trends and Future Directions in the School Closure Literature." *Educational Researcher* 51 (1): 58–65. https://doi.org/10.3102/0013189x211050944.

Fahle, Erin M., Thomas J. Kane, Tyler Patterson, Sean F. Reardon, Douglas O. Staiger, and Elizabeth A. Stuart. 2023. "School District and Community Factors Associated with Learning Loss During the COVID-19 Pandemic." Cambridge, MA.

Gallo, Marcos, Colin Camerer, Benjamin Manning, Katherine Milkman, and Angela Duckworth. 2022a. "Zearn: Empathy Study," September. https://doi.org/10.17605/OSF.IO/KDP2U.

———. 2022b. "Zearn: Nudging Weekly Logins," September. https://doi.org/10.17605/OSF.IO/JFV3B.

Greene, Jeffrey A. 2022. "What Can Educational Psychology Learn From, and Contribute to, Theory Development Scholarship?" *Educational Psychology Review* 34 (4): 3011–35. https://doi.org/10.1007/s10648-022-09682-5.

Haleva, Livnat, Arnon Hershkovitz, and Michal Tabach. 2020. "Students' Activity in an Online Learning Environment for Mathematics: The Role of Thinking Levels." *Journal of Educational Computing Research* 59 (4): 686–712. https://doi.org/10.1177/0735633120972057.

Hanushek, Eric A. 2020. "Education Production Functions." *The Economics of Education*, 161–70. https://doi.org/10.1016/b978-0-12-815391-8.00013-6.

Helwig, Nathaniel E. 2022. "Ica: Independent Component Analysis." https://CRAN.R-project.org/package=ica.

Helwig, Nathaniel E., and Sungjin Hong. 2013. "A Critique of Tensor Probabilistic Independent Component Analysis: Implications and Recommendations for Multi-Subject fMRI Data Analysis." *Journal of Neuroscience Methods* 213 (2): 263–73. https://doi.org/10.1016/j.jneumeth.2012.12.009.

Hershcovits, Haviv, Dan Vilenchik, and Kobi Gal. 2020. "Modeling Engagement in Self-Directed Learning Systems Using Principal Component Analysis." *IEEE Transactions on Learning Technologies* 13 (1): 164–71. https://doi.org/10.1109/tlt.2019.2922902.

Hill, Heather C., Merrie L. Blunk, Charalambos Y. Charalambous, Jennifer M. Lewis, Geoffrey C. Phelps, Laurie Sleep, and Deborah Loewenberg Ball. 2008. "Mathematical Knowledge for Teaching and the Mathematical Quality of Instruction: An Exploratory Study." *Cognition and Instruction* 26 (4): 430–511. https://doi.org/10.1080/07370000802177235.

Honore, Bo E. 1992. "Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects." *Econometrica* 60 (3): 533. https://doi.org/10.2307/2951583.

Hyvärinen, A., and E. Oja. 2000. "Independent Component Analysis: Algorithms and Applications." *Neural Networks* 13 (4-5): 411–30. https://doi.org/10.1016/S0893-6080(00)00026-5.

James, W., and Charles Stein. 1992. "Estimation with Quadratic Loss." In, edited by Samuel Kotz and Norman L. Johnson, 443–60. New York, NY: Springer New York. http://link.

springer.com/10.1007/978-1-4612-0919-5_30.

Koch, Alexander, Julia Nafziger, and Helena Skyt Nielsen. 2015. "Behavioral Economics of Education." *Journal of Economic Behavior & Organization* 115 (July): 3–17. https://doi.org/10.1016/j.jebo.2014.09.005.

Lavecchia, A. M., H. Liu, and P. Oreopoulos. 2016. "Behavioral Economics of Education." *Handbook of the Economics of Education*, 1–74. https://doi.org/10.1016/b978-0-444-63459-7.00001-4.

Liljedahl, Peter, and Susan Oesterle. 2020. "Teacher Beliefs, Attitudes, and Self-Efficacy in Mathematics Education." *Encyclopedia of Mathematics Education*, 825–28. https://doi.org/10.1007/978-3-030-15789-0_149.

Meeter, Martijn. 2021. "Primary School Mathematics During the COVID-19 Pandemic: No Evidence of Learning Gaps in Adaptive Practicing Results." *Trends in Neuroscience and Education* 25 (December): 100163. https://doi.org/10.1016/j.tine.2021.100163.

Millo, Giovanni. 2017. "Robust Standard Error Estimators for Panel Models: A Unifying Approach." *Journal of Statistical Software* 82 (3). https://doi.org/10.18637/jss.v082.i03.

Qiu, Feiyue, Guodao Zhang, Xin Sheng, Lei Jiang, Lijia Zhu, Qifeng Xiang, Bo Jiang, and Ping-kuo Chen. 2022. "Predicting Students' Performance in e-Learning Using Learning Process and Behaviour Data." *Scientific Reports* 12 (1). https://doi.org/10.1038/s41598-021-03867-8.

Ran, Hua, Murat Kasli, and Walter G. Secada. 2020. "A Meta-Analysis on Computer Technology Intervention Effects on Mathematics Achievement for Low-Performing Students in K-12 Classrooms." *Journal of Educational Computing Research* 59 (1): 119–53. https://doi.org/10.1177/0735633120952063.

Ran, Hua, Nam Ju Kim, and Walter G. Secada. 2021. "A Meta-Analysis on the Effects of Technology's Functions and Roles on Students' Mathematics Achievement in K-12 Classrooms." *Journal of Computer Assisted Learning* 38 (1): 258–84. https://doi.org/10.1111/jcal.12611.

Sadaf, Ayesha, Tong Wu, and Florence Martin. 2021. "Cognitive Presence in Online Learning: A Systematic Review of Empirical Research from 2000 to 2019." *Computers and Education Open* 2 (December): 100050. https://doi.org/10.1016/j.caeo.2021.100050.

Salazar, Addisson, Arturo Serrano, and Luis Vergara. 2007. "Non-Parametric ICA Reveals Learning Styles in Education Activities Through the Web." *2007 IEEE Workshop on Machine Learning for Signal Processing*, August. https://doi.org/10.1109/mlsp.2007.4414316.

Shin, Dongjo, and Jaekwoun Shim. 2020. "A Systematic Review on Data Mining for Mathematics and Science Education." *International Journal of Science and Mathematics Education* 19 (4): 639–59. https://doi.org/10.1007/s10763-020-10085-7.

StataCorp. 2023. *Stata Statistical Software*. College Station, TX: StataCorp LLC.

Wang, Xiaofang Sarah, Laura B. Perry, Anabela Malpique, and Tobias Ide. 2023. "Factors Predicting Mathematics Achievement in PISA: A Systematic Review." *Large-Scale Assessments in Education* 11 (1): 24. https://doi.org/10.1186/s40536-023-00174-8.

Young, Jamaal Rashad. 2016. "Unpacking TPACK in Mathematics Education Research: A Systematic Review of Meta-Analyses." *International Journal of Educational Methodology*

2 (1): 19–29. https://doi.org/10.12973/ijem.2.1.19.
"Zearn Math: Top-Rated Math Learning Platform." 2023. https://about.zearn.org/.
Zierer, Klaus. 2021. "Effects of Pandemic-Related School Closures on Pupils' Performance and Learning in Selected Countries: A Rapid Review." *Education Sciences* 11 (6): 252. https://doi.org/10.3390/educsci11060252.

Supplementary Information

Supplementary Methods

Independent Component Analysis (ICA)

We implement the FastICA algorithm (Hyvärinen and Oja 2000) to estimate independent components from our dataset. In this model, matrix $X = \{x_{ij}\}_{I \times J}$, consisting of I samples across J random variables, is expressed as a linear mixture of independent components C, represented by:

$$X = C'M + E.$$

Here, C holds the independent components, M is a mixing matrix, and E denotes the noise. The aim is to minimize mutual information between components in C, which is achieved by maximizing their marginal negentropy, thereby rendering the columns of C statistically independent.

The FastICA process begins by transforming X into a whitened matrix Y, ensuring uncorrelated variables with unit variance. This transformation is achieved through eigenvalue decomposition:

$$Y = X \times \left[\frac{\mathbf{v_1}}{\sqrt{\lambda_1}} \quad \frac{\mathbf{v_2}}{\sqrt{\lambda_2}} \quad \frac{\mathbf{v_3}}{\sqrt{\lambda_3}} \right],$$

where $(\lambda_1, \lambda_2, \lambda_3)$ and $(\mathbf{v_1}, \mathbf{v_2}, \mathbf{v_3})$ are, respectively, the eigenvalues and eigenvectors of $\frac{X'X}{I}$.

Afterward, the algorithm approximates the negentropy with

$$\hat{\theta}(c_n) = (\mathbb{E}[\ln(\cosh(c_n))] - \mathbb{E}[\ln(\cosh(z))])^2$$

where $c_n, n \in 1, 2, 3$, is one of the components, and z is a Gaussian variable with zero mean and unit variance. FastICA iteratively maximizes this value across all components, producing an orthogonal rotation matrix $R_{3\times3}$ such that $C = YR$.

For more details on ICA and the FastICA algorithm, see Hyvärinen and Oja (2000) and Helwig and Hong (2013).

Table 5: **?(caption)**

	[,1]	[,2]	[,3]
User.Session	0.4165322719	0.451370148	0.001295111
RD.elementary_schedule	0.0076554444	0.024375608	0.306049645
RD.whole_group_fluency	-0.0063450741	0.603612018	0.184724978
RD.mission_overview	0.0445833820	0.368615508	0.176495098
Guided.Practice.Completed	0.6518680472	0.066611541	-0.025873879
Tower.Completed	0.5580778557	0.137609474	-0.020545776
Fluency.Completed	0.8461717525	-0.013370087	0.012011877
Number.Gym.Activity.Completed	0.6552718233	-0.022244610	0.020881909
RD.grade_level_overview	0.0864051200	0.159300387	0.376527633
Tower.Stage.Failed	0.8957480772	-0.033264621	0.024505982
RD.optional_problem_sets	0.0223275294	0.656082120	-0.044115969
RD.student_notes_and_exit_tickets	0.1022378252	0.453560027	-0.035846444
Kindergarten.Activity.Completed	0.0435167903	-0.010110749	0.026577726
Tower.Struggled	0.8708113137	-0.056438084	0.021328851
RD.k_mission	0.0032089431	-0.043084861	0.237268044
RD.small_group_lessons	0.0182925416	0.718673789	0.074374561
RD.whole_group_word_problems	0.0168869349	0.650144552	0.126945331
RD.teaching_and_learning_approach	0.0122530967	-0.023189022	0.339360854
RD.optional_homework	0.0230270789	0.600875832	-0.048874965
RD.k_schedule	0.0045675377	-0.049101845	0.290690089
RD.assessments	0.0004377112	0.524428851	0.094042126
RD.assessments_answer_key	-0.0129605691	0.413379007	0.155801984
RD.pd_course_notes	-0.0077436060	0.077386100	0.637619258
RD.curriculum_map	0.0025866282	-0.006693604	0.314113884
RD.pd_course_guide	0.0020074670	-0.022928976	0.643445920

Censored Panel Regression

We can't really use fixed effects because there does not exist a sufficient statistic allowing the fixed effects to be conditioned out of the likelihood (Honore 1992).

Fixed-Effects Regression

Hausman Test

Supplementary Tables

Marginal Effects

Table 6: **?(caption)**

```
Error in tcrossprod(summary(ica_fe_model)$coefficients[-1, "Estimate"], : non-conformable arg
```

Supplementary Discussion

Robustnes Checks

- Additional Analyses: Detailed list of supplementary analyses to be conducted.

Supplementary Equations

Supplementary Notes