

# Discovering Data-Driven Nudges to Help Students Learn More Math

Marcos Gallo      Sean Hu      Ben Manning  
Angela Duckworth      Colin Camerer

## Abstract

The COVID-19 pandemic has exacerbated learning deficits, particularly in mathematics [13]. Adaptive online learning platforms offer a promising strategy to mitigate these adverse effects [36], but designing effective interventions to improve student outcomes via such platforms remains a challenge. Here, we present a two-phase study that uses granular teacher and student engagement data from a large online Math platform in the United States. First, using unsupervised learning techniques (Independent Component Analysis, ICA; and a modified Predicting Context Sensitivity (PCS) approach, drawn from Buyalskaya et al. [8], we identified critical teacher behaviors associated with improved student performance, namely empathy-focused engagement and strategic weekly login patterns. Building on these insights and in consultation with instructors, we developed two behaviorally-informed interventions: an empathy intervention encouraging teachers to view math problems from the student’s perspective, and a habit-building intervention emphasizing the importance of Friday logins for reflective analysis and proactive planning. In a large-scale randomized controlled trial ( $N = 140,461$  teachers across 22,281 schools), we demonstrated that the empathy intervention significantly increased student lesson completion by 4.09%. The habit condition also showed promise, increasing lesson completion by 1.81%, although this effect was not significantly different from the active control. Our approach demonstrates the potential of data-driven behavioral interventions to inform the development of more effective digital learning interventions, ultimately improving math education outcomes for students across diverse educational settings. Furthermore, we provide a generalizable framework for designing targeted interventions in various contexts where granular behavioral data are available.

## 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 [16]. Learning losses and disparities, particularly among younger students, have been significant due to reduced instructional time and remote learning challenges [53, 13]. Moreover, Ewing and Green [15] note that repeated school closures compounded these issues, emphasizing that the pandemic is not the only problem affecting math performance.

Emphasizing quality math instruction, particularly in pedagogical strategies, guidance, and teacher-student relations, is crucial for improving math performance, especially for students who have lost interest in mathematics or lack a sufficient foundation [47, 4]. Furthermore, teachers' deep mathematical knowledge is also important, enabling them to understand students' challenges better and provide adequate support [25, 4]. This combination of empathy and subject expertise highlights the importance of specialized training focusing on pedagogical skills and content knowledge.

The pandemic also accelerated the adoption of digital platforms for math education. Meeter [36] found that adaptive practice software effectively mitigated the adverse effects of school closures. This result is consistent with the qualitative evidence from Alabdulaziz [1], who reported that teachers found digital platforms beneficial in addressing the challenges of remote learning during the pandemic. Recent meta-analyses demonstrate that integrating digital tools and blended learning approaches improves student outcomes significantly [40, 39, 41]. Leveraging these tools and insights gained during the pandemic should help address longstanding educational challenges [15]. 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 [48, 7].

The increased use of digital platforms has provided more data to support the importance of student engagement in online learning. Blending online and in-person teaching methods can effectively enhance engagement and understanding, depending on the implementation [11, 30, 31, 35, 41, 49]. Teachers play a crucial role in helping students develop meta-cognitive skills to foster student engagement [20]. Furthermore, teachers' beliefs and self-efficacy toward technology integration influence their willingness to adopt innovative teaching practices [14, 32]. This relationship between student engagement and teacher attitudes suggests the importance of creating solutions that can be implemented across diverse educational settings. Our current study aims to address this need by developing and evaluating cost-effective, high-quality interventions designed to improve student outcomes through enhanced teacher engagement with the Zearn platform.

In this study, we partner with Zearn to address these topics. This math education platform reaches approximately 25% of elementary schools and over a million middle-school students in the United States. Zearn's approach integrates interactive digital lessons with hands-on teaching, aligning with the Common Core State Standards and providing a comprehensive educational experience [50].

Our study leverages this rich resource to offer an innovative approach to educational interventions using behavioral science principles. Focusing on quality of teaching, we align with Hanushek [21], who maintains that the efficacy of resource utilization supersedes quantity. We also follow current trends in providing cost-effective, easy-to-implement interventions (i.e., “nudges”) that offer engagement incentives for both teachers and students [34, 29, 28].

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 [42, 24, 38, 43, 44]. Subsequently, we aim to establish the causality of our interventions through a large-scale field experiment guided by recommendations from Greene [19] 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 Design

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 content is designed to promote intuitive understanding by progressing from concrete to pictorial to abstract examples. The platform offers a personalized experience, providing targeted remediation when a student encounters difficulties. Teachers can monitor individual student progress by tracking activities such as “Badges” and “Tower of Power.” Badges track student lesson completion, and the Tower of Power is an assessment that presents students with challenging problems at the end of each lesson. When students struggle or fail to answer correctly during a Tower of Power, the platform provides “Boosts” (i.e., hints or further explanations) and notifies teachers with a “Tower Alert.” These features aim to motivate students and provide valuable information for educators to support learning.

Other important variables from the dataset included teacher logins, file downloads, and specific interactions with educational content. Additionally, we accessed 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 in multiple times per week, exhibiting variations in the frequency and duration of their interactions. The student data, aggregated at the classroom-week level, showed a wide range of performance levels across various classrooms and schools. The standard deviations of key variables underscored this diversity, as detailed in the summary statistics of Table 1. This rich combination of teacher behaviors

and student performance metrics, carefully matched with a weekly frequency for each classroom, allowed for an in-depth analysis while upholding privacy standards. Note that the analyses in Study 1 underwent post hoc modifications to rectify some subsequently identified errors. These changes, while adjusting the coefficients slightly, did not significantly alter the overall results or patterns observed (refer to the SI for the original analyses).

Table 1: Study 1 Data. The table summarizes key educational metrics for Zearn teachers. The data was collected from July 2019 to June 2020.

	N	Aggregated Statistics per Teacher				
		Mean	Devia- tion	1st Quartile	Median	3rd Quartile
No. of teachers	2,506	-	-	-	-	-
No. of classes	4,228	1.69	0.97	1.0	1.0	2.0
No. of badges	5,230,607	2,087.23	1,627.26	1,035.2	1,682.8	2,643.2
Total minutes on Zearn	3,909,007	161.24	675.21	3.7	19.3	93.5

### Study 1a: Identifying Key Teacher Behaviors Associated with Student Performance Using Independent Component Analysis

Zearn offers a variety of activities for students and teachers, resulting in a large number of usage variables. We chose to summarize this large set of online activities into more easily interpretable variables through a dimensionality reduction technique. We employed Independent Component Analysis (ICA) on these teacher behavioral variables, given the non-Gaussian nature of the data [26]. This statistical approach allowed us to uncover latent variables that might have been obscured with traditional methods. The resulting components were weighted vectors of specific teacher activities and were estimated to maximize statistical independence in the data-generating process [26].

By examining the explained variance of each added component, we concluded that three independent components best portray the underlying dimensions of teacher behavior. (For more details, please refer to the SI section and Figure 2). It is important to note that the results presented here reflect a revised version of our pre-experimental analysis. Although we maintained the same analytical framework, we addressed several inaccuracies identified in a post hoc review (for detailed information, please refer to the Supplementary Information). These corrections yielded more precise and robust results.

The significant finding from the ICA was the prominence of the first Indepen-

dent Component (IC1), accounting for 15.39% of the variance in teacher Zearn activity, as indicated in Figure 1. This result is significant in nudge engineering, highlighting the need to focus interventions on elements encapsulated by IC1. Conversely, IC2 and IC3, with 12.56% and 6.3% variance explained, play lesser but still noteworthy roles.

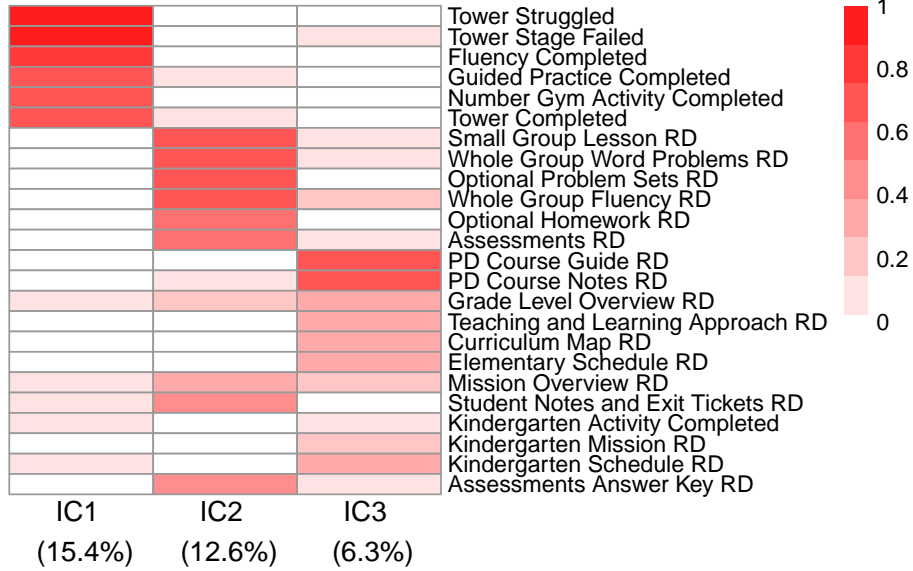


Figure 1: Independent Component Analysis (ICA) Results. The heatmap displays teacher actions in each row, while the columns represent the three independent components (IC1, IC2, IC3) that explained the most variance in the teacher behavioral data, with their respective percentage of variance explained 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. RD = Resource Download.

The activities with the highest weighting in IC1 included metrics associated with the “Tower of Power” (Struggled, Failed, and Completed) and other problem-solving exercises (i.e., Fluency exercises, Guided Practice, and Number Gym Activity). Notably, the top two variables in this component involved an interactive feature of the Tower of Power when students struggle to understand a concept (i.e., Tower Struggled or Failed). The platform offers tailored support through “Boosts,” which break down questions into smaller steps, helping students understand and correct their mistakes. Given that the metrics in the ICs pertain to teacher activity, IC1 suggests that teachers proactively engage with these problem-solving activities, sometimes deliberately making mistakes, to better understand the student experience. This interaction may help teachers devise strategies to break down complex problems into simpler steps, known as

instructional scaffolding [9, 5].

After consulting with Zearn educators and administrators (see SI for details), it was that teachers with high levels of IC1 apply an empathy-driven pedagogy. 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 (0.895), Fluency Completed (0.838), Guided Practice Completed (0.706), and Number Gym Activity Completed (0.661) were significantly higher compared to other activities. This pattern underlines this 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 a variety of Resource Downloads (RD), particularly Small Group Lessons (0.717) and Whole Group (i.e., entire class) Word Problems (0.669). We labeled this component as “Classroom Activities.” IC3, with strong weights on Resource Download (RD) for the Professional Development (PD) course guide (0.663) and course notes (0.657), was labeled “Professional Development.” This component, accounting for a smaller variance (6.296%), was harder to interpret due to more diverse activities.

Table 2: Regression of Log Badges on the Independent Components. The table displays the results of a panel regression model with clustered standard errors at the teacher level in parentheses.

	ln(Badges + 1)	
	All Schools	Zearn Curriculum
IC 1	0.010*(0.0043)	0.070***(0.0070)
IC 2	0.056***(0.0037)	0.056***(0.0040)
IC 3	-0.006***(0.0017)	-0.006***(0.0023)
No. of Classes	-0.067***(0.0079)	-0.038***(0.0100)
Average Intercept	1.276	1.133
R-squared	0.012	0.027
N	Teachers: 2506Classes: 4094Weeks: 39Total: 115532	Teachers: 1413Classes: 2389Weeks: 39Total: 67549
Note: * p<0.05; ** p<0.01; *** p<0.001		

Building upon these findings, we advanced to a fixed-effects panel regression model that accounted for various temporal and subject-specific variables and the potential impact of teachers handling multiple classes. 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 0.0101 ( $p = 0.0186$ ), suggesting a significant impact on student achievement. Upon the recommendation of Zearn administrators, a supplementary analysis was conducted, focused exclusively on schools using Zearn either as their core curriculum or a key supplementary resource, effectively excluding teachers who independently chose Zearn despite the absence of school-wide implementation. This subset analysis revealed an effect of 0.07 ( $p = 0$ ), indicating that IC1 is especially significant for those schools.

This correlation also has practical significance. It indicates that an increase in activities associated with IC1, such as increasing the encounters of ‘Tower Struggle’ by one standard deviation, is associated with an approximate 1.1 percent increase in student badges. Although this increase may seem small, it may be substantial in the context of nudges, where even modest changes often lead to far-reaching effects.

### **Study 1b: Identifying Key Teacher Habits Associated with Student Performance**

Prompted by Buyalskaya et al. [8], 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.
2. Average Minutes: Teachers’ weekly average time spent on the platform, measured in minutes.
3. Average Streak: Average number of consecutive days in which the teacher logs in, excluding weekends.
4. Average Weekday Streak: Average number of consecutive weekdays in which the teacher logs in (e.g., three Tuesday in a row).
5. Average Days Between Logins: Average number of days between two login instances.
6. Total Logins: Total number of teacher logins over the entire school year.

To account for any school-specific factors that may have influenced the relationship between teacher behavior and student achievement, we estimated a linear 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. Note that our regression omitted the login percentages from July and Sunday, periods with low login incidence, to avoid multicollinearity.

The regression results, as detailed in Table 3, revealed that all weekday login percentages positively affected student badges. However, Friday logins stood out significantly, suggesting that specific days of the week have more influence on habitual engagement.

Table 3: Regression of Log Average Badges on Habit Variables. The table displays the results of the regression model with standard errors clustered at the school level in parentheses. 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).

	Avg. Weekly Badges	
	Estimate (SE)	P-value
(Intercept)	4.513 (2.759)	.102
%Monday	0.859 (0.567)	.13
%Tuesday	0.725 (0.449)	.107
%Wednesday	0.866 (0.628)	.168
%Thursday	0.831 (0.515)	.107
%Friday	1.788 (0.507)	<.001
%Saturday	0.938 (1.167)	.421
%January	-4.118 (2.742)	.133
%February	-2.803 (2.843)	.324
%March	-3.818 (2.764)	.167
%April	-3.858 (2.789)	.167
%May	-3.193 (2.790)	.252
%June	-2.662 (2.965)	.369
%August	-4.008 (2.796)	.152
%September	-4.117 (2.783)	.139
%October	-3.757 (2.864)	.19
%November	-3.989 (2.785)	.152
%December	-3.185 (2.904)	.273
Avg. Minutes	0.011 (0.002)	<.001
Avg. Streak	0.061 (0.108)	.573
Avg. Days Between	0.002 (0.003)	.435
Logins		
Avg. Weekday Streak	-0.132 (0.185)	.475
Total Logins	0.011 (0.002)	<.001
R-squared	0.133	
N	4273	



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 9.68%. 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 13.78% in average weekly lesson completion. This outcome underscores the importance of strategic engagement rather than just login frequency.

Based on our data analysis and focus group discussions with teachers and administrators, we hypothesized that this pronounced effect may be due to two key factors. First, Friday logins could facilitate “Reflective Catch-Up,” enabling teachers to review and analyze the previous week’s activities and make necessary adjustments. Second, “Foresighted Planning” may occur 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 associated with 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 to increase student learning outcomes.

### **Study 2a: Using ICA Insights to Design an Empathy-Driven Teacher Intervention**

We used the first component from the ICA in Study 1a to design an “empathy” nudge. Our analysis in Study 1a indicated that this process was linked to the highest weighted behaviors in IC1 (note that the original analysis yielded an “empathy” coefficient of 0.0255 for the whole sample and 0.1434 for the curriculum-only subsample, versus 0.0101 and 0.07, respectively, in the revised analysis; refer to SI for details). 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. The emails contained key related messages that mentioned the general findings of a “recent analysis” (i.e., Study 1a), emphasized the importance of empathy in teaching math, and included 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: Using Habit Analysis Insights to Design a Friday Login Intervention

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, reminding them to log in on Fridays and suggesting reflective and planning behaviors to engage in during those sessions. These emails included testimonials from other teachers, research insights, and motivational messages to encourage habit formation. Teachers were also encouraged to review student progress and identify areas where students needed additional support (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.

## Intervention 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 marginal effects of the ‘Empathy’ and ‘Friday Login’ interventions compared to the study-wide control as a baseline. It also includes the results from the Friday-specific control group. We measure student achievement by the number of badges students earned.

Treatment	Total Badges	
	Estimate (SE)	P-value
Empathy	0.063 (0.029)	.0333
Friday	0.067 (0.029)	.0213
Friday Control	0.090 (0.029)	.00194

Table 4: Efficacy of Different Teacher Engagement Interventions on Student Learning Outcomes. The table showcases the marginal effects of the ‘Empathy’ and ‘Friday Login’ interventions compared to the study-wide control as a baseline. It also includes the results from the Friday-specific control group. We measure student achievement by the number of badges students earned.

Treatment	Total Badges	
	Estimate (SE)	P-value
Study Control	0.097 (0.015)	<.001
R-squared	0.685	
N	140461	

As stated in our pre-registration, we evaluated the impact of intervention messages on the number of math lessons completed by students during the four-week intervention period [18, 17]. 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 used the Benjamini-Hochberg (BH) procedure to compute adjusted p-values, which help to control for false discovery rates when conducting multiple comparisons [6]. 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 [27].

## 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 logins.

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 [25, 4]. In Study 1b, we discovered that teachers who logged into Zearn on Fridays had a notable impact on student math performance. This behavior suggested a commitment to continuous planning and support, echoing the findings of Blanchard et al. [7] 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. [14] 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 appears to have fostered a more engaging and supportive learning environment, a feature essential for the success of digital platforms [29, 28]. 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.

While yielding several insights, our study is limited by its focus on a specific demographic and educational context within Zearn. The data provide teacher and classroom interactions only within the online platform, which may reflect or influence in-class dynamics but does not directly measure in-class interactions. Consequently, generalizing our results to other educational settings, cultures, or age groups may be challenging. Future research could explore the relationship between online engagement patterns and in-class teaching practices. Further, we could not examine variation in performance among individual students within each classroom because of data aggregation, and future research could examine the variation in treatment effects among students within each classroom. 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 [3, 13]. 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 [10]. Future research could 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 dataset used in this study was automatically collected by Zearn’s servers during the 2019-2020 academic year (September 2019 to May 2020). The platform tracked user interactions, such as teacher and student logins, completed lessons by students, and professional development modules completed by teachers. Teacher actions were timestamped to the second, providing granular data on their behavior. To protect student privacy, student data was aggregated at the classroom-week level, including measures of student achievement and indicators of student struggles. We merged these data with a version of teacher data aggregated to the weekly level.

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.

To promote transparency and replicability of our study, we deposited the de-identified dataset and code used in our analyses in a publicly accessible database, available at the GitHub repository: [https://github.com/SeanHu0727/zearn\\_nudge.git](https://github.com/SeanHu0727/zearn_nudge.git)

### Inclusion Criteria

The dataset included various schools across Louisiana (see Fig S1 for geographic distribution). We also excluded inactive teachers (those with no recorded activity for over two months) from the dataset. We defined the following inclusion criteria for classrooms:

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

In study 1a, we further categorized Zearn usage into ‘curriculum’ and ‘non-curriculum’ cases. ‘Curriculum’ refers to scenarios where Zearn is integrated as a core component of the school’s daily schedule. ‘Non-curriculum’ cases, meanwhile, involve Zearn being used alongside different core curricula, resulting in varied consistency in usage.

## Analysis of 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 (i.e., the point at which adding more components yields diminishing increases in total explained variance), yielding three independent components. We used the `icafast` function from the R `ica` package for all ICAs conducted [22].

### Panel Regression

We estimated a fixed-effects panel regression model with the `plm` package [12] in R. The dependent variable was defined as:

$$\ln(\text{Badges}_{itc} + 1) = \beta_1 \text{IC1}_{it} + \beta_2 \text{IC2}_{it} + \beta_3 \text{IC3}_{it} \\ + \text{FE}_{\text{Teacher},i} + \text{FE}_{\text{Week},t} + \text{FE}_{\text{Class},c} + \epsilon_{itc}$$

where  $i$ ,  $c$ , and  $t$  index the teacher, class, and week, respectively. The model includes fixed effects for teacher ( $\text{FE}_{\text{Teacher},i}$ ), week ( $\text{FE}_{\text{Week},t}$ ), and class ( $\text{FE}_{\text{Class},c}$ ). Standard errors were clustered at the teacher level, calculated using the `vcovHC` function [37] in R.

## Study 1b

### Linear Regression

We estimated a linear model using the `lm` function [46] in R:

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

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 `vcovCL` function [51, 52] 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 [50] and was pre-registered 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 Preprocessing

Following the megastudy’s pre-registered analysis plan (<https://osf.io/dgpkn>), we restricted analyses to Zearn Math teachers who were assigned to one of the megastudy’s conditions and who taught in at least one classroom with at least one student. However, our analyses excluded teachers who: (1) did not receive any emails because they had inactive accounts, invalid email addresses, or had opted out of receiving messages ( $n = 133,722$  teachers), (2) neither logged onto the Zearn Math platform nor had an associated student who logged on the platform between March 1, 2021 and September 14, 2021 ( $n = 126,856$  teachers), (3) had more than 150 students associated with their Zearn Math account as of October 18, 2021 ( $n = 6,766$  teachers), or (4) had more than 6 classes associated with their Zearn Math account as of October 18, 2021 ( $n = 6,675$  teachers). Among Zearn Math classrooms associated with the remaining 149,097 teachers, we further excluded: (5) 9,143 classrooms associated with more than one Zearn Math teacher ( $n = 12,012$  teachers) and (6) 346 classrooms with grade levels corresponding to high school or post-high school ( $n = 141$  teachers).

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.



## Statistical Analysis

We followed our pre-registered 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 [18, 17]. We estimated a weighted ordinary least squares (OLS) regression with the `areg` command in Stata [45]. 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 Intervention Design and Implementation

We designed four emails encouraging teachers to adopt a more student-centered perspective when engaging with the Zearn platform to gain pedagogical content knowledge. The emails included testimonials from experienced teachers, research-based insights on the role of empathy in math education, and specific strategies for using Zearn’s features to understand and address student challenges. A population of 7,443 teachers was chosen at random to receive these emails.

## Study 2b: Friday Logins Intervention Design and Implementation

We designed four emails encouraging teachers to regularly log into the Zearn platform on Fridays. The emails emphasized the benefits of using Fridays for reflective review and proactive planning, highlighting how this practice could help teachers better support student learning. Teachers were provided with specific suggestions for activities during these Friday sessions, such as analyzing student progress data, identifying areas for improvement, and planning targeted interventions for the upcoming week. A total of 7,476 teachers was randomly selected to receive these emails. Additionally, to assess the unique impact of

the Friday login habit, we included an active control condition that received similar email prompts on Wednesdays with links to specific actions on Zearn but without a focus on Fridays. A group of 7,577 teachers were randomly chosen to participate in this control group.

## **Acknowledgments**

We thank Billy McRae, Michael Irvine, and Audrieanna Burgin for their valuable insights and support during our data analysis. We thank Zearn administrators and teachers for feedback on our intervention design. We also thank Cassandra Horri for the helpful comments that improved this manuscript.

## References

- [1] Mansour Saleh Alabdulaziz. “COVID-19 and the use of digital technology in mathematics education”. In: *Education and Information Technologies* 26.6 (June 28, 2021). 31 citations (Crossref) [2024-01-03], pp. 7609–7633. DOI: [10.1007/s10639-021-10602-3](https://doi.org/10.1007/s10639-021-10602-3).
- [2] M. Arellano. “PRACTITIONERS’ CORNER: Computing Robust Standard Errors for Within-groups Estimators \*”. In: *Oxford Bulletin of Economics and Statistics* 49.4 (Nov. 1987). 876 citations (Crossref) [2024-06-04], pp. 431–434. DOI: [10.1111/j.1468-0084.1987.mp49004006.x](https://doi.org/10.1111/j.1468-0084.1987.mp49004006.x). URL: <https://onlinelibrary.wiley.com/doi/10.1111/j.1468-0084.1987.mp49004006.x>.
- [3] A. V. Banerjee et al. “Remedying Education: Evidence from Two Randomized Experiments in India”. In: *The Quarterly Journal of Economics* 122.3 (Aug. 1, 2007). 599 citations (Crossref) [2024-01-03], pp. 1235–1264. DOI: [10.1162/qjec.122.3.1235](https://doi.org/10.1162/qjec.122.3.1235).
- [4] Dan Battey et al. “The interconnectedness of relational and content dimensions of quality instruction: Supportive teacher–student relationships in urban elementary mathematics classrooms”. In: *The Journal of Mathematical Behavior* 42 (June 2016). 18 citations (Crossref) [2024-01-03], pp. 1–19. DOI: [10.1016/j.jmathb.2016.01.001](https://doi.org/10.1016/j.jmathb.2016.01.001). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0732312316300013>.
- [5] Penny L. Beed, E. Marie Hawkins, and Cathy M. Roller. “Moving Learners toward Independence: The Power of Scaffolded Instruction”. In: *The Reading Teacher* 44.9 (1991). Publisher: [Wiley, International Reading Association], pp. 648–655. URL: <http://www.jstor.org/stable/20200767>.
- [6] Yoav Benjamini and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1 (Jan. 1995). 22907 citations (Crossref) [2024-01-03], pp. 289–300. DOI: [10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x). URL: <https://rss.onlinelibrary.wiley.com/doi/10.1111/j.2517-6161.1995.tb02031.x>.
- [7] Margaret R. Blanchard et al. “Investigating Technology-Enhanced Teacher Professional Development in Rural, High-Poverty Middle Schools”. In: *Educational Researcher* 45.3 (Apr. 2016). 57 citations (Crossref) [2024-01-03], pp. 207–220. DOI: [10.3102/0013189x16644602](https://doi.org/10.3102/0013189x16644602).
- [8] Anastasia Buyalskaya et al. “What can machine learning teach us about habit formation? Evidence from exercise and hygiene”. In: *Proceedings of the National Academy of Sciences* 120.17 (Apr. 25, 2023), e2216115120. DOI: [10.1073/pnas.2216115120](https://doi.org/10.1073/pnas.2216115120). URL: <https://pnas.org/doi/10.1073/pnas.2216115120>.
- [9] Zhihui Cai et al. “Effects of Scaffolding in Digital Game-Based Learning on Student’s Achievement: a Three-Level Meta-analysis”. In: *Educational Psychology Review* 34.2 (June 2022). 35 citations (Crossref) [2024-06-04], pp. 537–574. DOI: [10.1007/s10648-021-09655-0](https://doi.org/10.1007/s10648-021-09655-0). URL: <https://link.springer.com/10.1007/s10648-021-09655-0>.

- [10] Alan C.K. Cheung and Robert E. Slavin. “The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis”. In: *Educational Research Review* 9 (June 2013). 383 citations (Crossref) [2024-01-03], pp. 88–113. DOI: [10.1016/j.edurev.2013.01.001](https://doi.org/10.1016/j.edurev.2013.01.001).
- [11] Po-Jui Chiang, Ya-Wen Lin, and Chih-Lung Tseng. “The Effect of Blended Learning in Mathematics Course”. In: *EURASIA Journal of Mathematics, Science and Technology Education* 13.3 (Dec. 9, 2016). 27 citations (Crossref) [2024-01-03]. DOI: [10.12973/eurasia.2017.00641a](https://doi.org/10.12973/eurasia.2017.00641a).
- [12] Yves Croissant and Giovanni Millo. “Panel Data Econometrics in R : The **plm** Package”. In: *Journal of Statistical Software* 27.2 (2008). 561 citations (Crossref) [2023-12-30]. DOI: [10.18637/jss.v027.i02](https://doi.org/10.18637/jss.v027.i02). URL: <http://www.jstatsoft.org/v27/i02/>.
- [13] Giorgio Di Pietro. “The impact of Covid-19 on student achievement: Evidence from a recent meta-analysis”. In: *Educational Research Review* 39 (May 2023). 10 citations (Crossref) [2024-01-02], p. 100530. DOI: [10.1016/j.edurev.2023.100530](https://doi.org/10.1016/j.edurev.2023.100530). URL: <https://linkinghub.elsevier.com/retrieve/pii/S1747938X23000234>.
- [14] Peggy A. Ertmer et al. “Teacher beliefs and technology integration practices: A critical relationship”. In: *Computers & Education* 59.2 (Sept. 2012). 844 citations (Crossref) [2024-01-03], pp. 423–435. DOI: [10.1016/j.compedu.2012.02.001](https://doi.org/10.1016/j.compedu.2012.02.001).
- [15] Eve L. Ewing and Terrance L. Green. “Beyond the Headlines: Trends and Future Directions in the School Closure Literature”. In: *Educational Researcher* 51.1 (Oct. 13, 2021). 5 citations (Crossref) [2024-01-03], pp. 58–65. DOI: [10.3102/0013189x211050944](https://doi.org/10.3102/0013189x211050944).
- [16] Erin M. Fahle et al. *School district and community factors associated with learning loss during the COVID-19 pandemic*. Tech. rep. Cambridge, MA, May 2023.
- [17] Marcos Gallo et al. “Zearn: Empathy study”. In: (Sept. 30, 2022). Publisher: Open Science Framework. DOI: [10.17605/OSF.IO/KDP2U](https://doi.org/10.17605/OSF.IO/KDP2U). URL: <https://osf.io/kdp2u/>.
- [18] Marcos Gallo et al. “Zearn: Nudging weekly logins”. In: (Sept. 30, 2022). Publisher: Open Science Framework. DOI: [10.17605/OSF.IO/JFV3B](https://doi.org/10.17605/OSF.IO/JFV3B). URL: <https://osf.io/jfv3b/>.
- [19] Jeffrey A. Greene. “What Can Educational Psychology Learn From, and Contribute to, Theory Development Scholarship?” In: *Educational Psychology Review* 34.4 (May 23, 2022). 9 citations (Crossref) [2024-01-02], pp. 3011–3035. DOI: [10.1007/s10648-022-09682-5](https://doi.org/10.1007/s10648-022-09682-5). URL: <https://link.springer.com/10.1007/s10648-022-09682-5>.
- [20] Livnat Haleva, Arnon HersHKovitz, and Michal Tabach. “Students’ Activity in an Online Learning Environment for Mathematics: The Role of Thinking Levels”. In: *Journal of Educational Computing Research* 59.4 (Nov. 5, 2020). 7 citations (Crossref) [2024-01-03], pp. 686–712. DOI: [10.1177/0735633120972057](https://doi.org/10.1177/0735633120972057).

- [21] Eric A. Hanushek. “Education production functions”. In: *The Economics of Education* (2020). 29 citations (Crossref) [2024-01-03], pp. 161–170. DOI: [10.1016/b978-0-12-815391-8.00013-6](https://doi.org/10.1016/b978-0-12-815391-8.00013-6).
- [22] Nathaniel E. Helwig. *ica: Independent component analysis*. Tech. rep. 2022. URL: <https://CRAN.R-project.org/package=ica>.
- [23] Nathaniel E. Helwig and Sungjin Hong. “A critique of Tensor Probabilistic Independent Component Analysis: Implications and recommendations for multi-subject fMRI data analysis”. In: *Journal of Neuroscience Methods* 213.2 (Mar. 2013). 24 citations (Crossref) [2023-12-27], pp. 263–273. DOI: [10.1016/j.jneumeth.2012.12.009](https://doi.org/10.1016/j.jneumeth.2012.12.009). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0165027012004761>.
- [24] Haviv Hershcovits, Dan Vilenchik, and Kobi Gal. “Modeling Engagement in Self-Directed Learning Systems Using Principal Component Analysis”. In: *IEEE Transactions on Learning Technologies* 13.1 (Jan. 1, 2020). 8 citations (Crossref) [2024-01-03], pp. 164–171. DOI: [10.1109/tlt.2019.2922902](https://doi.org/10.1109/tlt.2019.2922902).
- [25] Heather C. Hill et al. “Mathematical Knowledge for Teaching and the Mathematical Quality of Instruction: An Exploratory Study”. In: *Cognition and Instruction* 26.4 (Sept. 26, 2008). 497 citations (Crossref) [2024-01-03], pp. 430–511. DOI: [10.1080/07370000802177235](https://doi.org/10.1080/07370000802177235).
- [26] A. Hyvärinen and E. Oja. “Independent component analysis: algorithms and applications”. In: *Neural Networks* 13.4-5 (June 2000). 5771 citations (Crossref) [2023-12-27], pp. 411–430. DOI: [10.1016/S0893-6080\(00\)00026-5](https://doi.org/10.1016/S0893-6080(00)00026-5). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0893608000000265>.
- [27] W. James and Charles Stein. “Estimation with Quadratic Loss”. In: ed. by Samuel Kotz and Norman L. Johnson. Series Title: Springer Series in Statistics DOI: [10.1007/978-1-4612-0919-5\\_30](https://doi.org/10.1007/978-1-4612-0919-5_30). New York, NY: Springer New York, 1992, pp. 443–460. URL: [http://link.springer.com/10.1007/978-1-4612-0919-5\\_30](http://link.springer.com/10.1007/978-1-4612-0919-5_30).
- [28] Alexander Koch, Julia Nafziger, and Helena Skyt Nielsen. “Behavioral economics of education”. In: *Journal of Economic Behavior & Organization* 115 (July 2015). 103 citations (Crossref) [2024-01-03], pp. 3–17. DOI: [10.1016/j.jebo.2014.09.005](https://doi.org/10.1016/j.jebo.2014.09.005).
- [29] A.M. Lavecchia, H. Liu, and P. Oreopoulos. “Behavioral Economics of Education”. In: *Handbook of the Economics of Education* (2016). 47 citations (Crossref) [2024-01-03], pp. 1–74. DOI: [10.1016/b978-0-444-63459-7.00001-4](https://doi.org/10.1016/b978-0-444-63459-7.00001-4).
- [30] Qing Li and Xin Ma. “A Meta-analysis of the Effects of Computer Technology on School Students’ Mathematics Learning”. In: *Educational Psychology Review* 22.3 (Sept. 2010). 265 citations (Crossref) [2024-06-04], pp. 215–243. DOI: [10.1007/s10648-010-9125-8](https://doi.org/10.1007/s10648-010-9125-8). URL: <http://link.springer.com/10.1007/s10648-010-9125-8>.
- [31] Shuqin Li and Weihua Wang. “Effect of blended learning on student performance in K-12 settings: A meta-analysis”. In: *Journal of Computer Assisted Learning* 38.5 (Oct. 2022). 12 citations (Crossref) [2024-06-04],

- pp. 1254–1272. DOI: [10.1111/jcal.12696](https://doi.org/10.1111/jcal.12696). URL: <https://onlinelibrary.wiley.com/doi/10.1111/jcal.12696>.
- [32] Peter Liljedahl and Susan Oesterle. “Teacher Beliefs, Attitudes, and Self-Efficacy in Mathematics Education”. In: *Encyclopedia of Mathematics Education* (2020). 2 citations (Crossref) [2024-01-03], pp. 825–828. DOI: [10.1007/978-3-030-15789-0\\_149](https://doi.org/10.1007/978-3-030-15789-0_149).
  - [33] J. Scott Long and Laurie H. Ervin. “Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model”. In: *The American Statistician* 54.3 (Aug. 2000). 299 citations (Crossref) [2024-06-04], pp. 217–224. DOI: [10.1080/00031305.2000.10474549](https://doi.org/10.1080/00031305.2000.10474549). URL: <http://www.tandfonline.com/doi/abs/10.1080/00031305.2000.10474549>.
  - [34] Kathleen Lynch et al. “Strengthening the Research Base That Informs STEM Instructional Improvement Efforts: A Meta-Analysis”. In: *Educational Evaluation and Policy Analysis* 41.3 (Sept. 2019). 54 citations (Crossref) [2024-06-04], pp. 260–293. DOI: [10.3102/0162373719849044](https://doi.org/10.3102/0162373719849044). URL: <http://journals.sagepub.com/doi/10.3102/0162373719849044>.
  - [35] Dona N. Mawardi, Chatarina A. Budiningsih, and Sugiman. “Blended Learning Effect on Mathematical Skills: A Meta-Analysis Study”. In: *Ingénierie des systèmes d’information* 28.1 (Feb. 28, 2023). 0 citations (Crossref) [2024-06-04], pp. 197–204. DOI: [10.18280/isi.280122](https://doi.org/10.18280/isi.280122). URL: <https://iicta.org/journals/isi/paper/10.18280/isi.280122>.
  - [36] Martijn Meeter. “Primary school mathematics during the COVID-19 pandemic: No evidence of learning gaps in adaptive practicing results”. In: *Trends in Neuroscience and Education* 25 (Dec. 2021). 31 citations (Crossref) [2024-01-03], p. 100163. DOI: [10.1016/j.tine.2021.100163](https://doi.org/10.1016/j.tine.2021.100163).
  - [37] Giovanni Millo. “Robust Standard Error Estimators for Panel Models: A Unifying Approach”. In: *Journal of Statistical Software* 82.3 (2017). 45 citations (Crossref) [2023-12-30]. DOI: [10.18637/jss.v082.i03](https://doi.org/10.18637/jss.v082.i03). URL: <http://www.jstatsoft.org/v82/i03/>.
  - [38] Feiyue Qiu et al. “Predicting students’ performance in e-learning using learning process and behaviour data”. In: *Scientific Reports* 12.1 (Jan. 10, 2022). 35 citations (Crossref) [2024-01-03]. DOI: [10.1038/s41598-021-03867-8](https://doi.org/10.1038/s41598-021-03867-8).
  - [39] Hua Ran, Murat Kasli, and Walter G. Secada. “A Meta-Analysis on Computer Technology Intervention Effects on Mathematics Achievement for Low-Performing Students in K-12 Classrooms”. In: *Journal of Educational Computing Research* 59.1 (Sept. 10, 2020). 21 citations (Crossref) [2024-01-03], pp. 119–153. DOI: [10.1177/0735633120952063](https://doi.org/10.1177/0735633120952063). URL: <http://journals.sagepub.com/doi/10.1177/0735633120952063>.
  - [40] Hua Ran, Nam Ju Kim, and Walter G. Secada. “A meta-analysis on the effects of technology’s functions and roles on students’ mathematics achievement in K-12 classrooms”. In: *Journal of Computer Assisted Learning* 38.1 (Sept. 29, 2021). 11 citations (Crossref) [2024-01-03], pp. 258–284. DOI: [10.1111/jcal.12611](https://doi.org/10.1111/jcal.12611). URL: <https://onlinelibrary.wiley.com/doi/10.1111/jcal.12611>.

- [41] Ayesha Sadaf, Tong Wu, and Florence Martin. “Cognitive Presence in Online Learning: A Systematic Review of Empirical Research from 2000 to 2019”. In: *Computers and Education Open* 2 (Dec. 2021). 14 citations (Crossref) [2024-01-03], p. 100050. DOI: [10.1016/j.caeo.2021.100050](https://doi.org/10.1016/j.caeo.2021.100050). URL: <https://linkinghub.elsevier.com/retrieve/pii/S2666557321000215>.
- [42] Addison Salazar, Arturo Serrano, and Luis Vergara. “Non-Parametric ICA reveals Learning Styles in Education Activities through the Web”. In: *2007 IEEE Workshop on Machine Learning for Signal Processing* (Aug. 2007). 0 citations (Crossref) [2024-01-03]. DOI: [10.1109/mlsp.2007.4414316](https://doi.org/10.1109/mlsp.2007.4414316).
- [43] Raghad Al-Shabandar et al. “Analyzing Learners Behavior in MOOCs: An Examination of Performance and Motivation Using a Data-Driven Approach”. In: *IEEE Access* 6 (2018). 29 citations (Crossref) [2024-01-03], pp. 73669–73685. DOI: [10.1109/access.2018.2876755](https://doi.org/10.1109/access.2018.2876755).
- [44] Dongjo Shin and Jaekwoun Shim. “A Systematic Review on Data Mining for Mathematics and Science Education”. In: *International Journal of Science and Mathematics Education* 19.4 (May 14, 2020). 22 citations (Crossref) [2024-01-03], pp. 639–659. DOI: [10.1007/s10763-020-10085-7](https://doi.org/10.1007/s10763-020-10085-7).
- [45] StataCorp. *Stata Statistical Software*. College Station, TX: StataCorp LLC, 2023.
- [46] R Core Team. *R: A language and environment for statistical computing*. Tech. rep. Vienna, Austria, 2023. URL: <https://www.R-project.org/>.
- [47] Xiaofang Sarah Wang et al. “Factors predicting mathematics achievement in PISA: a systematic review”. In: *Large-scale Assessments in Education* 11.1 (June 20, 2023). 2 citations (Crossref) [2024-01-02], p. 24. DOI: [10.1186/s40536-023-00174-8](https://doi.org/10.1186/s40536-023-00174-8). URL: <https://largescaleassessmentsineducation.springeropen.com/articles/10.1186/s40536-023-00174-8>.
- [48] Jamaal Rashad Young. “Unpacking TPACK in Mathematics Education Research: A Systematic Review of Meta-Analyses”. In: *International Journal of Educational Methodology* 2.1 (Nov. 15, 2016). 15 citations (Crossref) [2024-01-03], pp. 19–29. DOI: [10.12973/ijem.2.1.19](https://doi.org/10.12973/ijem.2.1.19).
- [49] Qing Yu et al. “Effectiveness of blended learning on students’ learning performance: a meta-analysis”. In: *Journal of Research on Technology in Education* (Oct. 11, 2023), pp. 1–22. DOI: [10.1080/15391523.2023.2264984](https://doi.org/10.1080/15391523.2023.2264984). URL: <https://www.tandfonline.com/doi/full/10.1080/15391523.2023.2264984>.
- [50] *Zearn Math: Top-Rated Math Learning Platform*. 2023. URL: <https://about.zearn.org/>.
- [51] Achim Zeileis. “Econometric Computing with HC and HAC Covariance Matrix Estimators”. In: *Journal of Statistical Software* 11.10 (2004). 581 citations (Crossref) [2024-01-31]. DOI: [10.18637/jss.v011.i10](https://doi.org/10.18637/jss.v011.i10). URL: <http://www.jstatsoft.org/v11/i10/>.
- [52] Achim Zeileis, Susanne Köll, and Nathaniel Graham. “Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R”. In: *Journal of Statistical Software* 95.1 (2020). 232 citations (Cross-

ref) [2024-01-31]. DOI: [10.18637/jss.v095.i01](https://doi.org/10.18637/jss.v095.i01). URL: <http://www.jstatsoft.org/v95/i01/>.

- [53] Klaus Zierer. “Effects of Pandemic-Related School Closures on Pupils’ Performance and Learning in Selected Countries: A Rapid Review”. In: *Education Sciences* 11.6 (May 22, 2021). 57 citations (Crossref) [2024-01-03], p. 252. DOI: [10.3390/educsci11060252](https://doi.org/10.3390/educsci11060252).



## Supplementary Information

### Differences in the Original vs. Revised Analysis

As noted in the main text, we have corrected a few inaccuracies in our initial analysis, improving the robustness of our methodology. To be fully transparent, we have outlined below all the changes made from the original to the revised analysis. Please refer to Tables S3 to S5 in this supplementary document for the original results that inspired our treatments.

1. **Temporal Alignment Adjustments:** The original method used the standard week and year delineation based on the Gregorian calendar. The new analysis uses the ISO week date system to ensure correct week numbering, particularly around the transition from one year to the next, which corrected previous week misclassifications.
2. **Duplicate Record Management:** The dataset contains a small number of duplicate classroom-week pairs generated in classrooms with more than one teacher linked to it. The original method removed the first duplicate occurrence, consequently discarding data from teachers with larger ID numbers. The revised analysis orders and filters duplicate records by the number of classes each teacher manages, retaining data from teachers involved in fewer classes, thereby minimizing the inclusion of supervisory rather than direct instructional roles.
3. **Independent Component Analysis:** The original ICA used the `fastICA` package, presenting some inconsistencies: non-deterministic component sequencing, arbitrary sign inversions, and varied loading coefficients, even with a set seed. The updated approach adopts the `ica` package, ensuring orderly component arrangement and consistent outputs. Sign orientation is now standardized, maintaining the largest loading variable as positive. Further, for the models restricted to schools that use Zearn as a main component of their curriculum, the ICA was estimated separately (see Table S4).
4. **Model Selection Adjustment:** Initially, we estimated a random effects model. Following a Hausman test indicating inconsistency ( $\chi^2 = 81.31$ ,  $df = 4$ ,  $p < .001$ ), we transitioned to a fixed effects model. The revised approach also incorporates robust standard errors, adjusting for heteroskedasticity and autocorrelation [2, 33].
5. **Removed User.Session:** The original ICA included “User.Session.” This variable measures the frequency of a teacher’s logins to the Zearn platform. As such, it does not offer substantive insight into the pedagogical nature of the teachers’ interactions with the platform. We aimed to understand the educational impact of specific usage patterns rather than their frequency. Hence, “User.Session” was deemed a nuisance variable as it risked overshadowing more pertinent patterns related to instructional engagement.

and effective pedagogical strategies. In particular, this variable mainly loads onto the “empathy” component. Its exclusion in our revised analysis has confirmed that the identified patterns genuinely reflect empathy in teaching approaches (see Tables S3 and S4).

## Supplementary Methods

### Independent Component Analysis (ICA)

We implement the FastICA algorithm [26] 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:

$$Y = X \times \begin{bmatrix} \frac{\mathbf{v}_1}{\sqrt{\lambda_1}} & \frac{\mathbf{v}_2}{\sqrt{\lambda_2}} & \frac{\mathbf{v}_3}{\sqrt{\lambda_3}} \end{bmatrix},$$

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  $R_{3 \times 3}$  such that  $C = YR$ .

For more details on ICA and the FastICA algorithm, see Hyvärinen and Oja [26] and Helwig and Hong [23].

### Focus Group Discussions

In 2021, we conducted regular meetings (once a month, on average) with Zearn employees to discuss the interpretation of our data analyses. Additionally, on

April 6th and 13th, 2021, we held “office hours” with Zearn teachers and administrators. For these specific meeting we prepared the following questions, although due to time limitations, we were unable to ask them systematically:

1. With what regularity do you find logging into Zearn most helpful?
2. What tasks do you typically do on Zearn?
3. Please order the tasks you mentioned in how important they are in helping your teaching (from most important to least important).
4. Please explain why [top option from question 3] is the most important.
5. Have you done activities and exercises designed for the students on Zearn?
6. If so, with what regularity do you do activities and exercises designed for the students on Zearn?

Below are some of the most relevant quotes from these meetings:

“We could layer in some of the insights here, [for example], prompting teachers to go and do towers and get towers and go through mediation paths. I think that could be something we layer into one of those A/B tests.” - B.M., Zearn administrator

“If students are not successful in Tower of Power, they get stuck, and it is the loop of self defeat - I know what I need to do, but I can’t do it. Teachers don’t check the tower alerts report enough so they can defeat that cycle.” - Math coach from CA

“For everything I do, I try to look through the kid’s eyes. I look for the lightbulb moment, and the thing that finally gets kids to understand it.” -5th grade teacher from IL

“One thing is when kids really understand something and have that “aha” moment - and you know you have gotten through. And that could be academically or socially.” - Middle school teacher from CA

“It took my students 50-60 minutes to get through a lesson. I have some kids who get fatigued by the length of the digital instruction and program. The boosts may not help them. Need to orient them to the Zearn lesson.” - Zearn Teacher

Table 5: Independent Component Analysis (ICA) Results. This table displays the weights of each teacher activity in the three independent components (ICs). Notably, these metrics pertain to teacher activity on the platform, not student actions.

	IC 1	IC 2	IC 3
Tower Struggled	0.89	-0.02	0.02
Tower Stage Failed	0.89	-0.01	0.02
Fluency Completed	0.84	0.00	0.00
Guided Practice Completed	0.71	0.04	-0.02
Number Gym Activity Completed	0.66	0.00	0.00
Tower Completed	0.65	0.09	0.00
RD.grade_level_overview	0.12	0.16	0.35
RD.student_notes_and_exit_tickets	0.10	0.43	-0.04
Kindergarten Activity Completed	0.06	-0.01	0.03
RD.mission_overview	0.03	0.37	0.17
RD.k_schedule	0.03	-0.05	0.28
RD.teaching_and_learning_approach	0.02	-0.02	0.33
RD.optional_problem_sets	0.01	0.65	-0.05
RD.optional_homework	0.01	0.60	-0.05
RD.whole_group_word_problems	0.01	0.67	0.10
RD.elementary_schedule	0.00	0.02	0.28
RD.small_group_lessons	0.00	0.72	0.06
RD.curriculum_map	0.00	0.00	0.31
RD.k_mission	0.00	-0.03	0.23
RD.assessments	-0.01	0.55	0.08
RD.pd_course_guide	-0.01	-0.03	0.66
RD.whole_group_fluency	-0.01	0.64	0.15
RD.assessments_answer_key	-0.01	0.45	0.14
RD.pd_course_notes	-0.01	0.08	0.66

Table 6: Marginal Effects of ICA Components on Badges. Marginal effects are calculated by multiplying the ICA coefficients by the mean of each component and dividing by the standard deviation of the corresponding variable.

	Marginal Effect of 1 SD
Tower Struggled	4.94047241
Tower Stage Failed	4.88767850
Fluency Completed	4.73596883
Guided Practice Completed	4.09623322
Number Gym Activity Completed	3.70304928
Tower Completed	3.66105471
RD.student_notes_and_exit_tickets	0.58765178
Kindergarten Activity Completed	0.16896987
RD.optional_homework	0.04351804
RD.optional_problem_sets	0.01554424
RD.small_group_lessons	-0.77319686
RD.assessments	-0.89092131
RD.whole_group_word_problems	-1.04695424
RD.mission_overview	-1.18494365
RD.assessments_answer_key	-1.24335142
RD.whole_group_fluency	-1.44767299
RD.k_mission	-1.53853195
RD.k_schedule	-1.70513738
RD.grade_level_overview	-1.78839364
RD.elementary_schedule	-1.87201584
RD.teaching_and_learning_approach	-2.08801621
RD.curriculum_map	-2.09292751
RD.pd_course_guide	-4.46985513
RD.pd_course_notes	-4.50891505
	Marginal Effect of 1 Unit
Tower Struggled	7.13716807
Tower Stage Failed	2.03274465
Fluency Completed	3.47202255
Guided Practice Completed	2.69186933
Number Gym Activity Completed	5.62688906
Tower Completed	2.73406190
RD.student_notes_and_exit_tickets	0.87619609
Kindergarten Activity Completed	0.73256635
RD.optional_homework	0.06412106
RD.optional_problem_sets	0.01590469
RD.small_group_lessons	-0.81204265
RD.assessments	-1.75942328
RD.whole_group_word_problems	-1.70447661
RD.mission_overview	-2.03205446
RD.assessments_answer_key	-3.71051193
RD.whole_group_fluency	-2.84089609
RD.k_mission	-5.42947068
RD.k_schedule	-84.55016610
RD.grade_level_overview	-13.00956584
RD.elementary_schedule	-34.91100596
RD.teaching_and_learning_approach	-41.00397400
RD.curriculum_map	-46.13048544
RD.pd_course_guide	-67.42936802
RD.pd_course_notes	-31.16319161

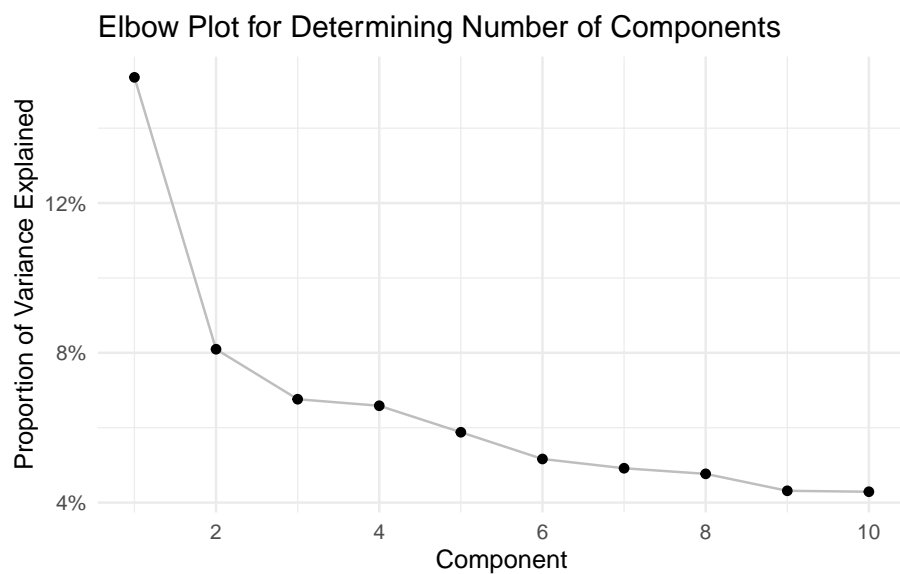


Figure 2: Elbow (Scree) Plot for Determining Optimal Number of Components. The plot displays the proportion of variance explained by each independent component (IC). The optimal number of components is indicated by the ‘elbow’ of the plot, where the variance explained by each additional component is minimal.

### Number of Teachers by Parish in Louisiana

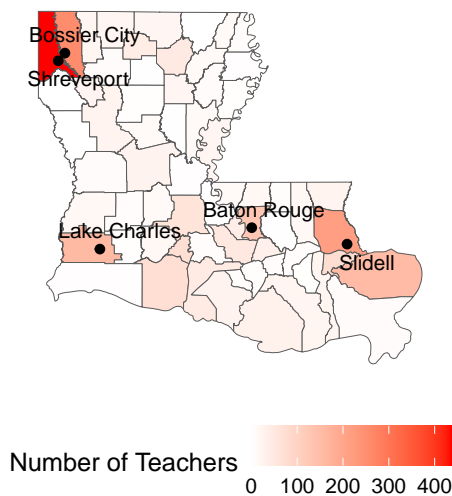


Figure 3: Geographical distribution of teachers across various parishes in Louisiana, and the top 5 cities with the highest number of teachers.

Table 7: Independent Component Analysis (ICA) Results without including User Sessions.

	IC 1	IC 2	IC 3
Tower Struggled	0.89	-0.03	0.02
Tower Stage Failed	0.88	-0.02	0.03
Fluency Completed	0.83	0.00	0.00
Guided Practice Completed	0.71	0.04	-0.02
Number Gym Activity Completed	0.66	-0.01	0.01
Tower Completed	0.65	0.09	0.00
User Session	0.54	0.38	-0.01
RD.grade_level_overview	0.11	0.16	0.36
RD.student_notes_and_exit_tickets	0.11	0.43	-0.04
Kindergarten Activity Completed	0.06	-0.01	0.03
RD.mission_overview	0.03	0.37	0.18
RD.k_schedule	0.03	-0.05	0.28
RD.teaching_and_learning_approach	0.02	-0.02	0.32
RD.optional_problem_sets	0.02	0.65	-0.05
RD.optional_homework	0.01	0.60	-0.05
RD.small_group_lessons	0.01	0.72	0.06
RD.elementary_schedule	0.00	0.02	0.28
RD.curriculum_map	0.00	0.00	0.31
RD.whole_group_word_problems	0.00	0.66	0.11
RD.k_mission	0.00	-0.03	0.23
RD.pd_course_guide	0.00	-0.02	0.66
RD.assessments	-0.01	0.54	0.09
RD.pd_course_notes	-0.01	0.07	0.65
RD.whole_group_fluency	-0.02	0.63	0.17
RD.assessments_answer_key	-0.02	0.44	0.15
[1]	0.15750742	0.12489921	0.06064235

**PCS Analysis**

**Supplementary Tables**

**Supplementary Figures**

**Supplementary Discussion**

**Original Analysis**

**Robustness Checks**