Predicting Repeated Behavior in Behavioral Sciences

Applying Reinforcement Learning in Teacher Decision-Making

Marcos Gallo

Table of contents

1 Introduction

1.1 Predicting Repeated Behavior in the Behavioral Sciences

Applying Reinforcement Learning in Teacher Decision-Making

1.2 A Novel Approach

2 Context and Research Questions

2.1 The Zearn Platform

Zearn is an online math-teaching platform.

- Model the decision-making of teachers
- Understand how they adapt their teaching strategies to optimize student achievement.

2.2 Data - Zearn Platform

Personalized learning experience for students. Teachers track student progress and make informed decisions.

- **Classroom structure:** Self-paced online lessons and small group instruction.
- **Badge system for student achievement:** Students earn badges upon completing lessons (mastery of specific skills). Track student progress and motivate them to continue learning.
- **Tower Alerts:** Real-time notifications sent to teachers when a student struggles with a specific concept. Teachers can provide support and address learning gaps.
- **Teacher selection and criteria:** Consistently use the platform and work in traditional school settings.
- **Variables of interest:** Teacher effort, student performance, lesson completion, and the time spent by both teachers and students on the platform.

2.3 Research Questions

3 Theory

3.1 Teacher effort and student achievement

3.1.1 Education production function

3.1.2 Context and experience

3.2 Reinforcement Learning to Capture Patterns in Repeated Behavior

3.2.1 Why Reinforcement Learning?

- RL is inspired by the way animals learn from their experiences
- An agent in RL represents a decision-maker
- Actions: choices made by the agent
- Environment: the context in which the agent makes decisions
- Observations: information the agent receives about the environment
- Rewards: feedback received by the agent based on the actions taken

In the context of predicting repeated behavior, RL algorithms can be used to model the decision-making process of individuals or groups, such as teachers, by learning from the patterns in their actions and the resulting outcomes.

RL: an **agent** learns to make decisions by interacting with an **environment**. Through trial and error, agent learns the best **actions** to take in different situations to achieve its **goals**.

$$\text{Agent} \xrightarrow[\text{Actions}]{\text{Performs}} \text{Environment} \xrightarrow[\text{Observations, Rewards}]{\text{Provides}} \text{Agent}$$

Suitability for modeling teacher decision-making

- Captures the dynamic and sequential nature of teaching
- Example: $s_t = (TowerAlerts_t, RD\_resources\_t)$, $a_t = (RD\_small\_group\_lessons\_t, TimeSpent\_t)$
- Allows for the exploration of optimal teaching strategies in response to students' progress and engagement
- Example: Balancing between focusing on struggling students and challenging high-performing students

Assumptions, objective functions, and tradeoffs

Assumes teachers make decisions to maximize long-term rewards (e.g., student learning outcomes)

Objective: $\max_\pi \mathbb{E}[\sum_{t=0}^{T} \gamma^t r(s_t, a_t)|\pi]$

Balances the tradeoffs between exploration (trying new teaching strategies) and exploitation (using known effective strategies)

Example: ϵ-greedy strategy

Flexibility and robustness

- Adapts to changes in the learning environment and individual student needs
- Example: Adapting to new curriculum or varying levels of student preparedness
- Allows for the incorporation of various state, action, and reward variables
- Example: Including external factors such as school policies or testing schedules
- Can be tailored to different educational contexts and objectives
- Example: Customizing the model for different grade levels
- Tradeoff between learning (exploring) and optimizing (exploiting)

Example:

- State: Current progress of students in the class.
- Actions: Assigning additional practice, providing personalized feedback, adjusting lesson plans.
- Rewards: Improved student performance, student engagement, reduced learning gaps.

$$State(S) \xrightarrow[\text{Action (A)}]{\text{Teacher Decides}} NewState(S') \xrightarrow[\text{Reward (R)}]{\text{Resulting Outcome}} \text{Feedback}$$

3.2.2 Q-Learning Model

3.2.3 Actor-Critic Model

1. Full RL framework
2. Policy learning and state value learning
3. Eligibility traces for delayed rewards

3.2.4 Gaussian Policy Model

D. Continuous Control through a Gaussian Policy 1. Probability distribution over actions

3.3 RL in teaching and education

1. Markov decision processes
2. Instructional actions, objectives, and costs

4 Data

4.1 Descriptive Statistics and Visualizations

The data represents various aspects of Zearn schools including identifiers, usage data, and demographic information. The dataset contains information for 22153 classrooms and 13921 teachers, with an average of 20 students per classroom. Various transformations and computations were performed on the data to prepare it for analysis, including calculating the number of distinct teachers, total students, and total weeks per school.

Descriptive statistics were computed for different measures such as the number of unique teachers, total students, and total weeks per school. Proportions were also calculated for various variables like poverty and income. The resulting statistics and proportions are displayed in Table 1 and Table 2, respectively. Refer to Table 3 for detailed information on the summary statistics for different variables by grade level.

The geographical distribution of teachers across Louisiana and the top 5 cities with the highest number of teachers are presented in Figure 1.

Table 1: Summary statistics for the schools

	Mean	Standard Deviation	Minimum	Maximum
Teachers	12.08	11.84	1	72
Students_Total	268.72	279.49	1	3,289
Weeks_Total	9.30	7.16	1	39

Table 2: Proportions of different variables.

	Proportions
Poverty Level	
0-40% (Low)	20.12%

40-75% (Mid-High)	53.44%
75%+ (High)	20.52%
Income	
$1–27,999	5.12%
$28K-31,999	5.81%
$32K-34,999	7.39%
$35K-36,999	4.61%
$37K-38,999	3.52%
$39K-40,999	2.55%
$41K-42,999	5.34%
$43K-44,999	3.66%
$45K-47,999	7.11%
$48K-51,999	18.48%
$52K-54,999	4.16%
$55K-59,999	2.95%
$60K-64,999	9.1%
$65K-69,999	4.02%
$70K-80,999	8.8%
$81K-93,999	2.94%
$94K+	1.62%
Other	
Charter_Schools	7.01%
Schools_with_Paid_Account	71.04%

```
|                                                                    |   0%
|                                                                    |   1%
|=                                                                   |   1%
|=                                                                   |   2%
|==                                                                  |   2%
|==                                                                  |   3%
|===                                                                 |   4%
|===                                                                 |   5%
|====                                                                |   5%
```

```
|
|====                                                             |   6%
|
|=====                                                            |   6%
|
|=====                                                            |   7%
|
|=====                                                            |   8%
|
|======                                                           |   8%
|
|======                                                           |   9%
|
|=======                                                          |  10%
|
|========                                                         |  11%
|
|========                                                         |  12%
|
|=========                                                        |  12%
|
|=========                                                        |  13%
|
|==========                                                       |  14%
|
|==========                                                       |  15%
|
|===========                                                      |  15%
|
|===========                                                      |  16%
|
|============                                                     |  17%
|
|============                                                     |  18%
|
|=============                                                    |  19%
|
|==============                                                   |  19%
|
|==============                                                   |  20%
|
|===============                                                  |  22%
|
```

```
|================                                                 |  23%
|
|=================                                                |  24%
|
|=================                                                |  25%
|
|==================                                               |  26%
|
|====================                                             |  28%
|
|====================                                             |  29%
|
|=====================                                            |  30%
|
|======================                                           |  32%
|
|========================                                         |  34%
|
|=========================                                        |  36%
|
|==========================                                       |  37%
|
|===========================                                      |  40%
|
|============================                                     |  42%
|
|==============================                                   |  44%
|
|================================                                 |  47%
|
|==================================                               |  50%
|
|===================================                              |  51%
|
|====================================                             |  54%
|
|=====================================                            |  58%
|
|=====================================                            |  58%
|
|=========================================                        |  62%
|
|==========================================                       |  65%
```

```
|
|=================================================                     |  69%
|
|==================================================                    |  72%
|
|====================================================                  |  77%
|
|======================================================                |  79%
|
|=======================================================               |  80%
|
|=========================================================             |  85%
|
|===========================================================           |  86%
|
|=============================================================         |  92%
|
|===============================================================       |  94%
|
|======================================================================| 100%
```

4.2 Unit of Analysis: Classroom-Week

4.2.1 Zearn's Eye View

Time-series data for teacher effort and student achievement A. Weekly aggregation of data B. Privacy concerns for student data

4.2.2 Measuring Student Achievement

4.2.3 Change in Active Users and Badges per Active User Over Time

The average number of active students over time is shown in Figure 2, and the total number of student logins over time is illustrated in Figure 3.

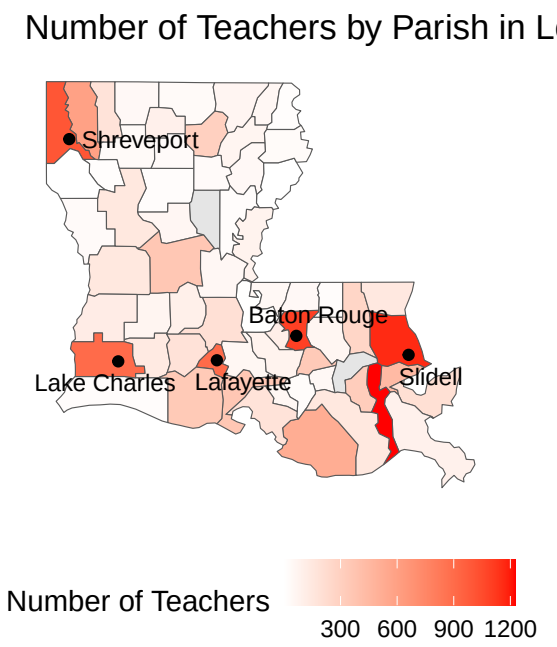

Figure 1: Geographical distribution of teachers across various parishes in Louisiana, and the top 5 cities with the highest number of teachers.

Table 3: **?(caption)**

(a)

Grade Level	Sessions per Student	Minutes per Student	Badges per Student	Tower A
Overall, N = 262,771	4.68 (3.34)	2.96 (1.98)	0.76 (0.68)	
Kindergarten, N = 16,746	3.61 (2.77)	2.37 (1.82)	1.02 (0.90)	
1st, N = 41,224	4.63 (2.98)	2.93 (1.99)	0.79 (0.68)	
2nd, N = 48,428	4.63 (3.01)	2.97 (1.99)	0.77 (0.67)	
3rd, N = 50,958	4.71 (3.14)	3.05 (1.97)	0.78 (0.66)	
4th, N = 50,803	4.85 (3.53)	3.07 (1.96)	0.71 (0.64)	
5th, N = 49,569	5.02 (3.98)	3.08 (1.95)	0.72 (0.63)	
6th, N = 3,006	3.06 (2.28)	1.81 (1.98)	0.35 (0.56)	
7th, N = 1,035	3.27 (2.22)	2.15 (1.94)	0.41 (0.54)	
8th, N = 1,002	2.66 (1.63)	1.64 (1.82)	0.26 (0.47)	

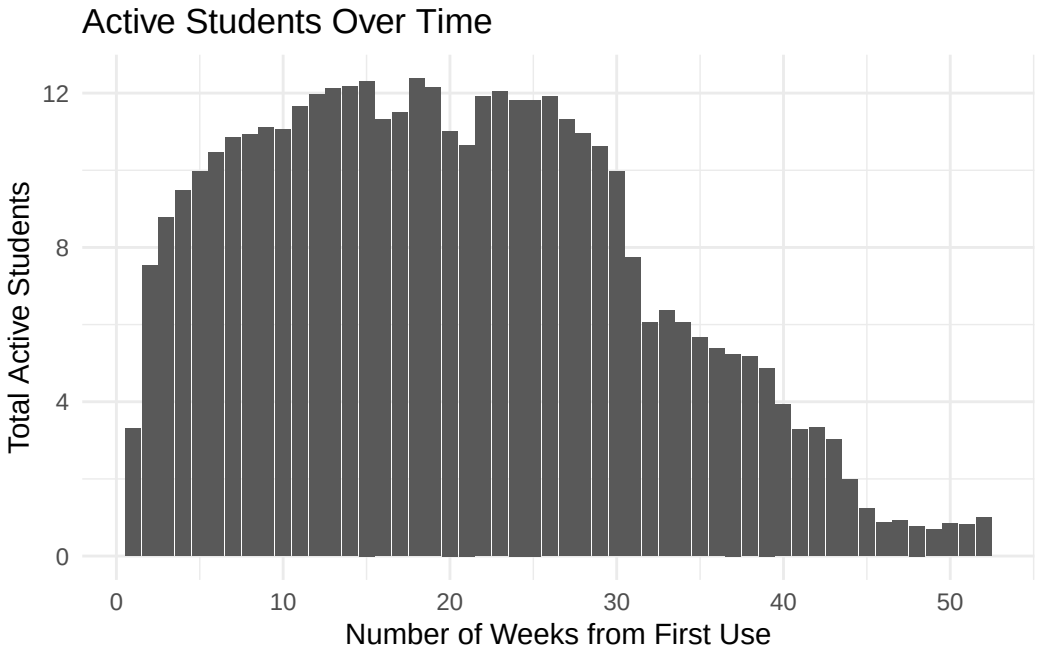

Figure 2: Change in the average number of active students over time.

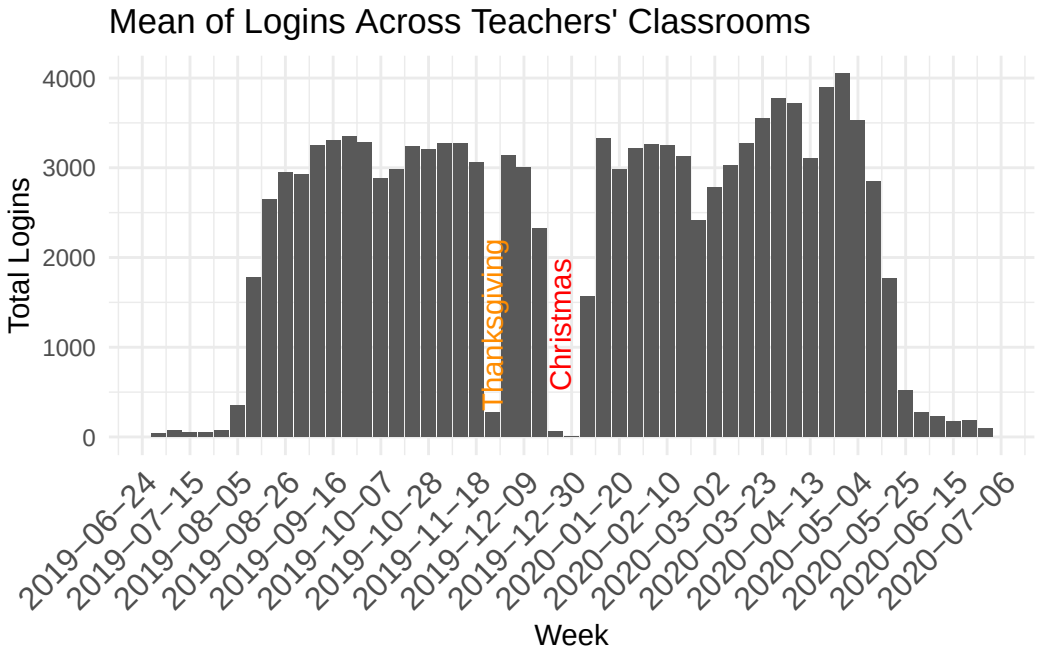

Figure 3: Total number of student logins over time.

4.3 Exclusion criteria

1. Traditional schools and consistent platform usage
2. Criteria for inclusion in the study We remove teachers with more than 4 classrooms and those who have logged in for less than 12 weeks in total. We exclude classrooms in the 6th to 8th grades, as those are a small proportion of our dataset.

4.4 Variables of interest

Teacher log-ins and time spent on the platform Student lesson completion (badges) D. State variables 1. Tower Alerts 2. Student time usage

4.4.1 Visualizing Relationships Between Variables

- Use correlation analysis to find relationships between variables
- Identify variables that may be strong predictors for the reinforcement learning model

Figure 4 displays the correlation matrix of selected variables.

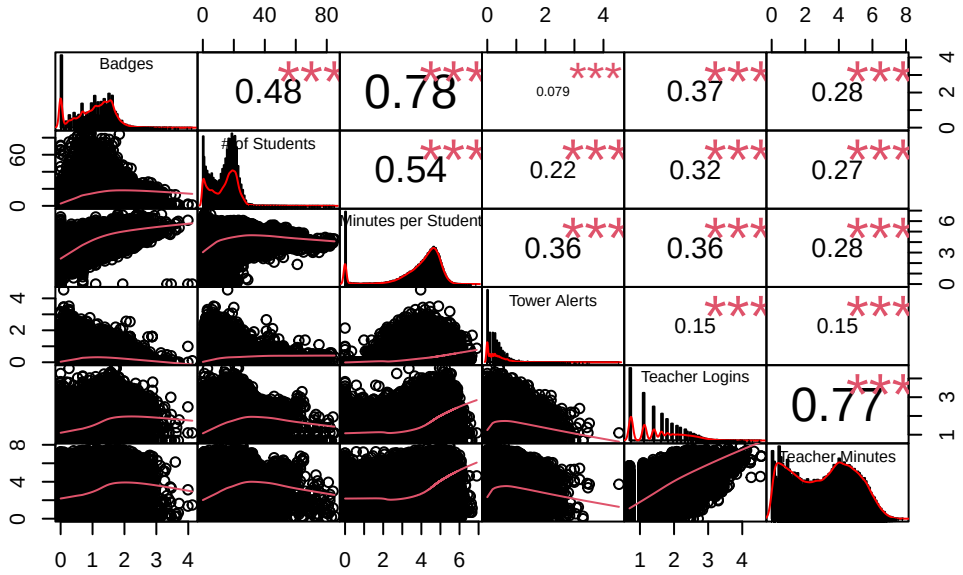

Figure 4: This graph represents the correlation between variables after log transformation

4.4.1.1 Are Some Badges Harder than Other?

4.4.2 Dimensionality Reduction

In order to capture the choices and trade-offs that teachers make, we used a Principal Component Analysis (PCA). This approach was taken to condense the multifaceted nature of our variables: Teacher Minutes, Teacher Sessions, and a series of data points including resources downloaded.

PCA was performed to mitigate the high-dimensionality of the dataset, seeking to encapsulate the maximum statistical information. We then calculated the correlations between the "Badges per Student" and the following variables: Teacher Minutes, and the three principal components (PC1, PC2, PC3). This facet of the analysis, which had a unit of analysis at the teacher level, was conducted to analyze the relationship between badges and the selected variables for each Teacher.

?@fig-pca displays the Scree plot showing the optimal number of principal components.

The Non-negative Matrix Factorization (NMF) operates as follows:

The original matrix can be seen as a detailed description of all the teachers' behaviors. Each row in the matrix represents a unique teacher, and each column represents a specific behavior or action the teacher might take. The entry in a specific row and column then corresponds to the occurrence, frequency, or intensity of that behavior for that particular teacher. After the NMF, we have two matrices:

1. **Basis Matrix (W)**: This matrix represents underlying behavior patterns. Each column can be seen as a "meta-behavior" or a group of behaviors that tend to occur together. It is an abstraction or summary of the original behaviors.
2. **Mixture Matrix (H)**: This matrix shows the extent to which each "meta-behavior" is present in each teacher. Each entry in this matrix represents the contribution of a "meta-behavior" to a particular teacher's behaviors.

By looking at these matrices, we can identify underlying patterns of behaviors (from the basis matrix) and see how these patterns are mixed and matched in different teachers (from the mixture matrix). This can be a powerful way of summarizing and interpreting complex behavioral data.

4.4.3 Interpreting Components

4.5 Connecting Variables to Reinforcement Learning Model

States

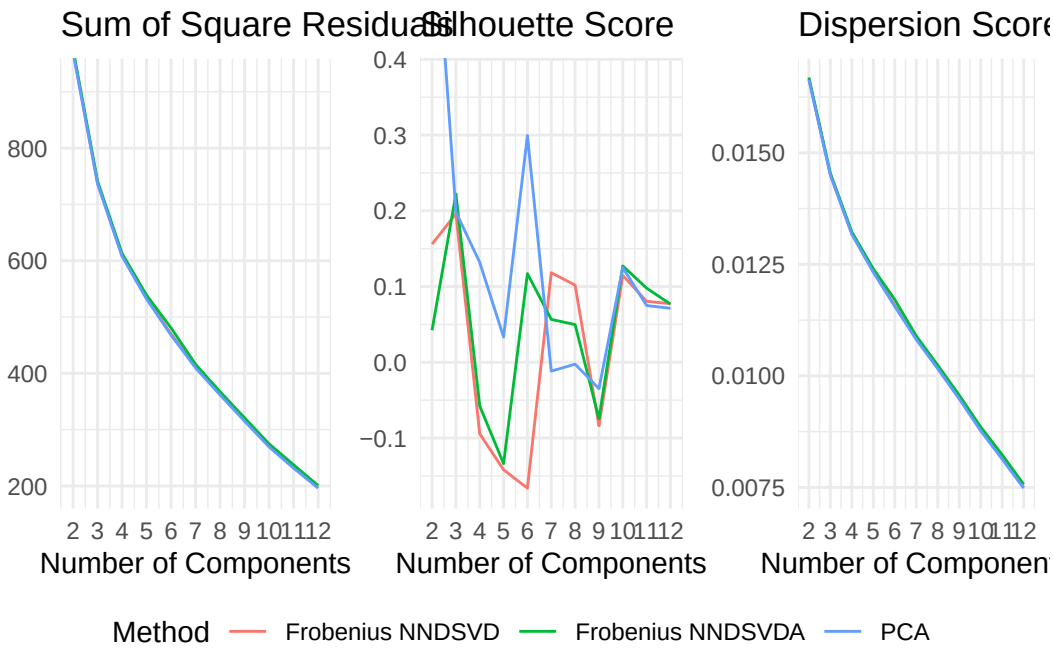

Figure 5: Comparison of residuals and silhouette scores for PCA, Frobenius, and Kullback-Leibler methods.

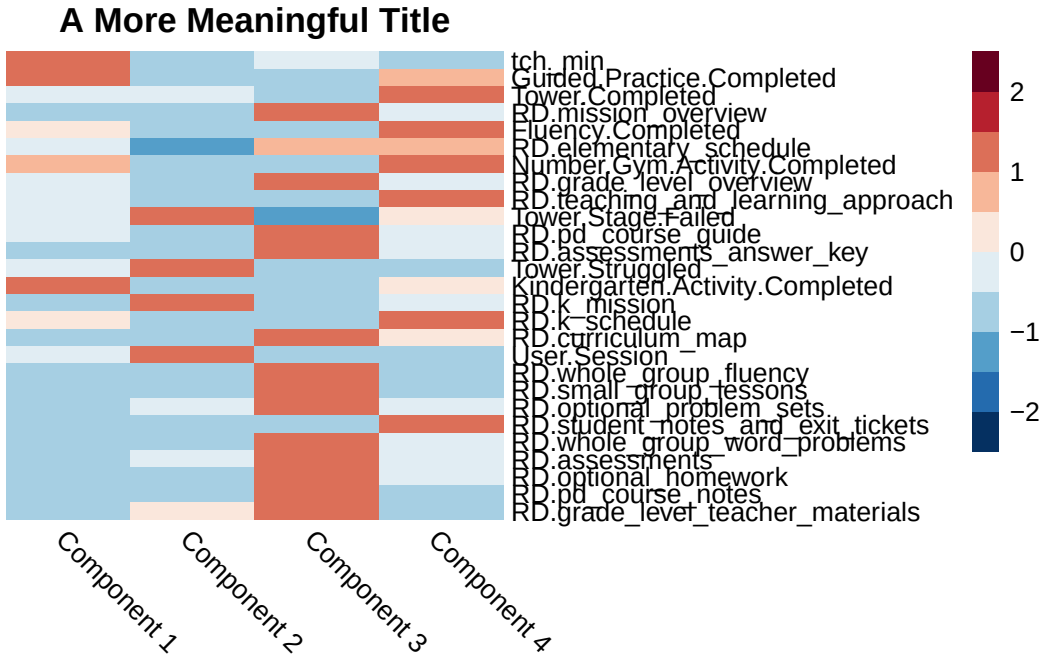

Figure 6: **?(caption)**

- Tower Alerts points to how many students are struggling with the content.
- Minutes per Student / Badges per Student
- Total Active Students
- Different combinations of these variables to create unique states.

Rewards

- Learning progress through objective measures such as badges, boosts.
- Quantify the effectiveness of teacher actions in promoting student learning.

Actions

- Teachers can download resources, engage in different teaching methods or activities.
- Teachers can choose how much time they spend online.
- RL optimizes the action selection based on the rewards observed.

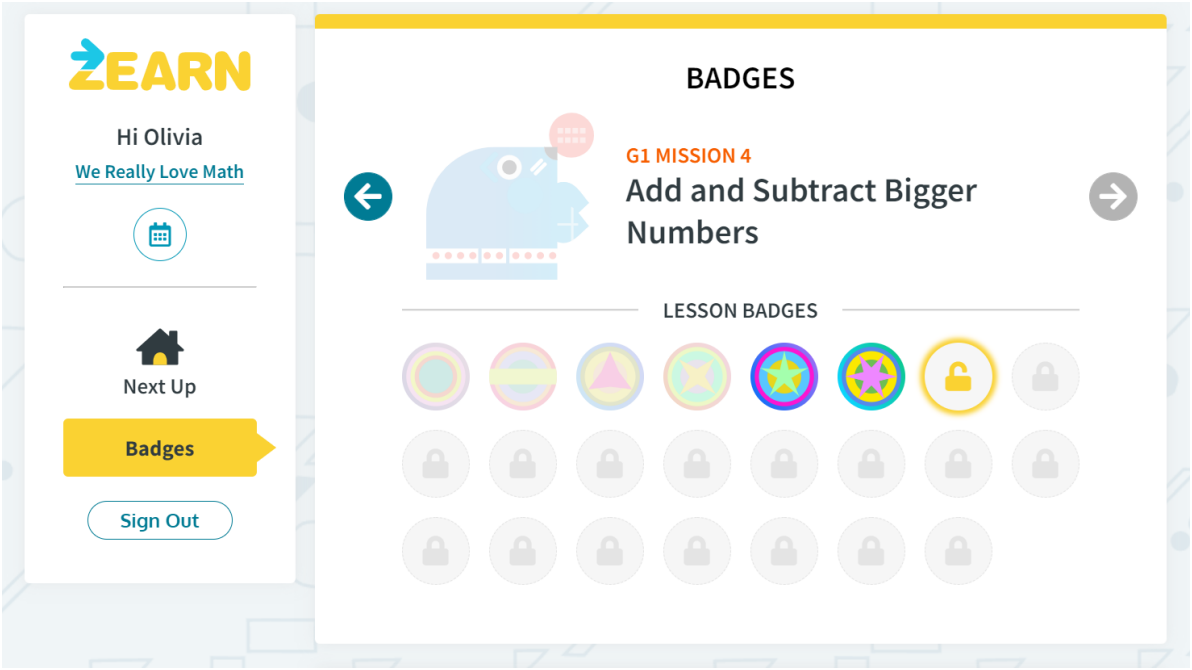

Image of the badge system for student achievement

5 Methods

5.1 Dynamic Analysis (Lau & Glimcher, 2005)

Introduction to dynamic analysis

- Uses response-by-response models to predict choice on each trial based on past reinforcers and choices
- Based on logistic regression, it captures the linear combination of past reinforcers and choices on each trial
- Flexible model incorporating effects of past reinforcers, choice history, and biases

Advantages of dynamic analysis in the context of Zearn dataset

- Captures the temporal dependencies and complex interactions between teacher actions, student outcomes, and learning environment
- Allows for the identification of optimal teaching strategies that evolve over time
- Enables the evaluation of the impact of various factors (e.g., curriculum, student engagement, etc.) on the decision-making process

Model Formulation

$$\log\left(\frac{p_{R,i}}{p_{L,i}}\right) = \sum_{j=1} \alpha_j (r_{R,i-j} - r_{L,i-j}) + \sum_{j=1} \beta_j (c_{R,i-j} - c_{L,i-j}) + \gamma,$$

Preliminary insights and findings

- Identification of key factors that influence teacher decision-making and student outcomes
- Evidence of adaptive teaching strategies that change in response to student progress and engagement
- Estimation of the relative impact of different teaching actions on student learning

5.2 Variable Selection

5.3 Q-learning Model

- Teacher Specific reward sensitivity
- Reward is a linear function of badges (reward sensitivity minus cost)

5.4 Actor-Critic Model

5.5 Gaussian Policy Model

5.6 Model Fit

5.6.1 Base Models: Random Effects Panel Logit

5.6.2 Hierarchical Bayesian Method

Bayesian updating using Stan software package Priors informed by grid search

5.7 Model Comparison

5.8 Heterogeneity

5.8.1 Across Teachers

5.8.2 Across Schools

5.8.3 Across Demographics

6 Results

6.1 Component Selection

6.2 Meta-Analysis Overall Results

Subsequently, we performed a meta-analysis of these correlations to reveal the pooled effect of our variables. In this case, we conducted a multivariate meta-analysis, offering the advantage of modeling multiple, potentially correlated, outcomes. We transformed the correlations using Fisher's z-transformation (to ensure a normal distribution of the correlations) and ran a random effects model with each unique combination of "Teacher" and "School".

The resulting multivariate meta-analysis provides a comprehensive estimate of the correlations for each outcome, considering the hierarchical structure of the data. Thus, we can understand the overarching relationships between the different outcomes and the Badges across diverse schools and teachers. This robust conclusion, therefore, provides a more resilient analysis than a simple correlation analysis.

?@fig-meta-analysis presents the results of a meta-analysis on the correlation.

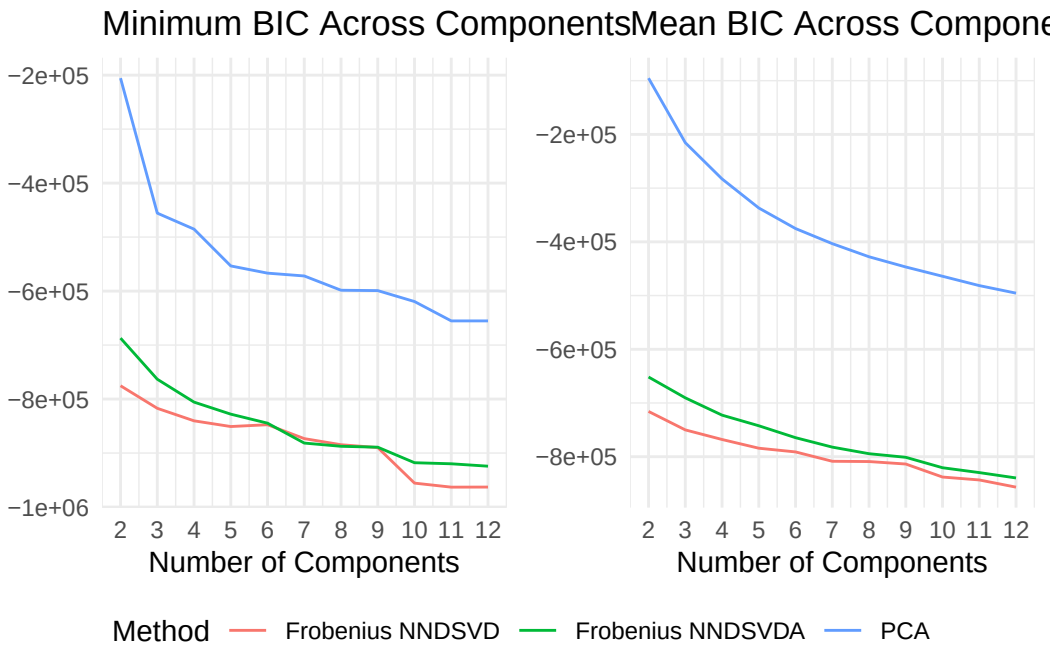

Figure 7: BICs

6.3 Base Models

In order to get a baseline understanding of the influence of our key variables on the number of badges per active user, we employed a series of panel data models, with control variables: 1) the number of classes each teacher is responsible for, 2) the grade level of the classes, and 3) the total number of students. The 'Minutes Model' considers the number of minutes each teacher spends on the platform. The 'PCA Models' incorporate the three principal components we derived earlier.

Both of these models use a random effects approach, which is suitable for our panel data structure and accounts for unobserved heterogeneity.

6.4 Models with Lags

Subsequently, we accounted for the temporal dynamics of our dataset by applying the Lau & Glimcher (2005) method. We introduced lagged variables into the models, thereby allowing us to account for temporal autocorrelation and potential delayed effects. We included lagged versions of the variables 'Teacher Minutes', 'PC1', 'PC2', 'PC3', and 'Badges per Students', with eight lags for each.

We then ran models with these lagged variables using the same random effects approach as in the base models.

In order to better understand and interpret the output of our models, we created a plot of the coefficients associated with each of the lagged variables. This plot allows us to see how the influence of each variable changes as the lag increases, and to compare these dynamics across variables.

6.5 Variable selection

To ensure that our models are parsimonious and to help determine which set of variables provides the best fit for the data, we compared the Bayesian Information Criterion (BIC) of the different models: the Minutes Model, PCA Model, and Lag Models. **?@tbl-bic** displays the BICs, with each row corresponding to a different model.

BIC penalizes models based on their complexity (number of parameters used) and the number of observations, favoring simpler models and models that fit the data better. In **?@tbl-bic**, the 8-lag PC1 Model has the lowest BIC value, suggesting that it provides the best fit to the data when considering both complexity and fit.

6.6 Q-Learning Analysis

6.7 Actor-Critic Analysis

6.8 Gaussian Policy

6.9 Model Comparison

6.10 Heterogeneity

7 Discussion

6. Comparing the performance of the models
7. Advantages and limitations of each model
8. Insights from each model

7.1 Implications for Teachers and Schools

3. Decision-making patterns
4. Optimal strategies
5. Implications for the education field
6. Potential impact on teaching practices
7. Policy recommendations

7.2 Limitations

3. Data limitations and biases
4. Model assumptions and simplifications
5. Generalizability of results

7.3 Challenges

7.4 Future research

6. Application to other educational contexts
7. Integration with other models and approaches
8. Expanding the scope of variables and data sources