

# Beyond Exclusion: The Role of High-Stake Testing on Attendance

Magdalena Bennett<sup>1</sup>

@maibennett  
m.bennett@austin.utexas.edu

Christopher Neilson<sup>2</sup> Nicolás Rojas<sup>3</sup>

<sup>1</sup> McCombs School of Business, The University of Texas at Austin

<sup>2</sup> Economics Department, Princeton University

<sup>3</sup> Teachers College, Columbia University

## Introduction

**Non-representative patterns of attendance** can skew how useful test scores measures are for accomplishing their goal. The **main objectives** of this paper are the following:

1. Understand the **average effect of testing on school attendance** across grades and performance
2. **Identify schools** that incentivize non-representative patterns of attendance by **combining causal inference methods and machine learning**
3. Help improve current **imputation methods**

## Methods

### Event Study

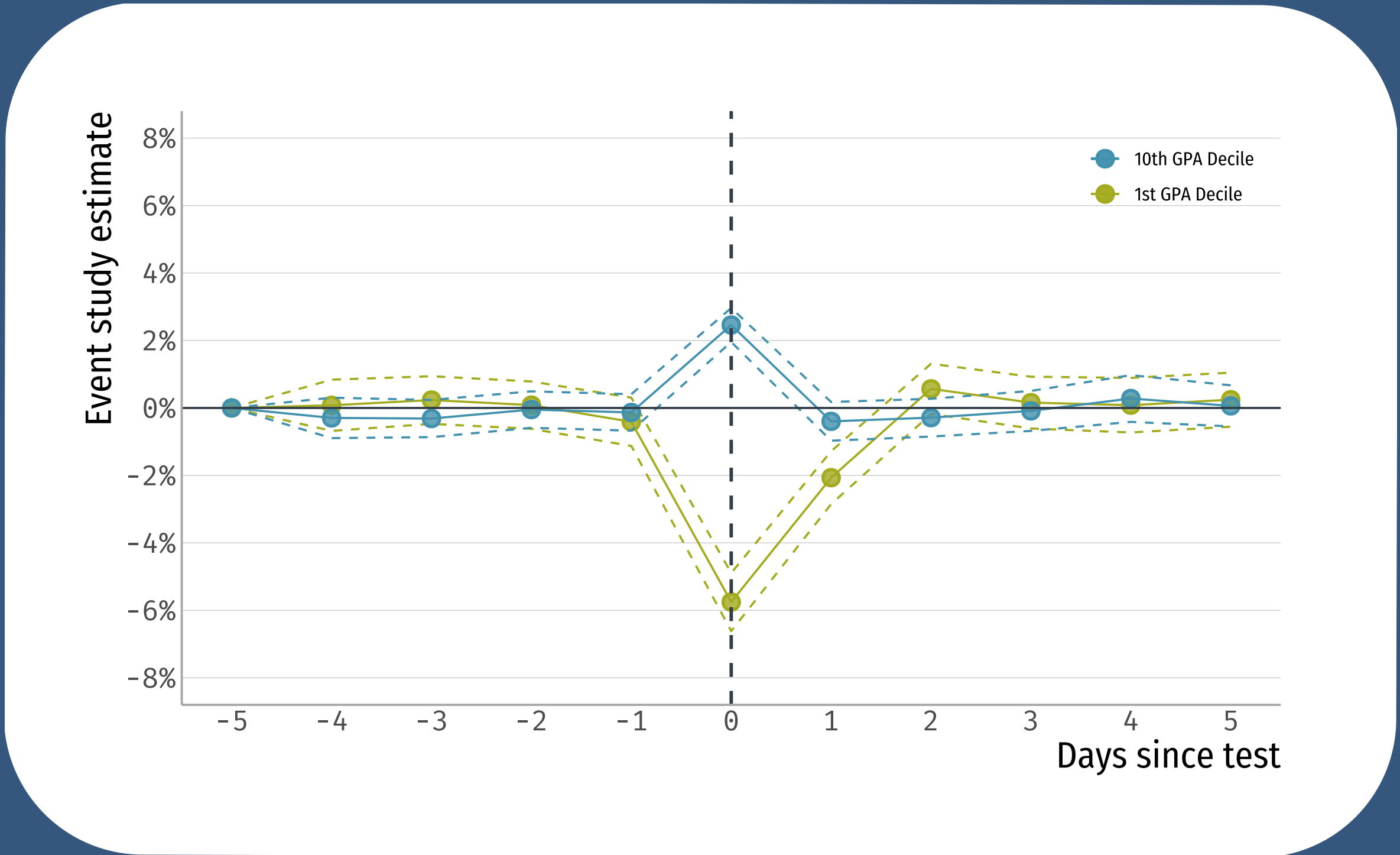
$$Y_{ipsgt} = \sum_{P=1}^5 \sum_{T=-4}^5 \tau^{PT} D_{ipsgt}^{PTG^*} + \gamma_{pt} + \alpha_i + \epsilon_{ipsgt}$$

- $Y_{ipsgt}$ : Attendance (1,0) for student  $i$ , from GPA group  $p$ , in school  $s$  and grade  $g$  for day  $t$ .
- $D_{ipsgt}^{PTG^*}$ : Indicator variable where  $G^*$  is the tested grade.

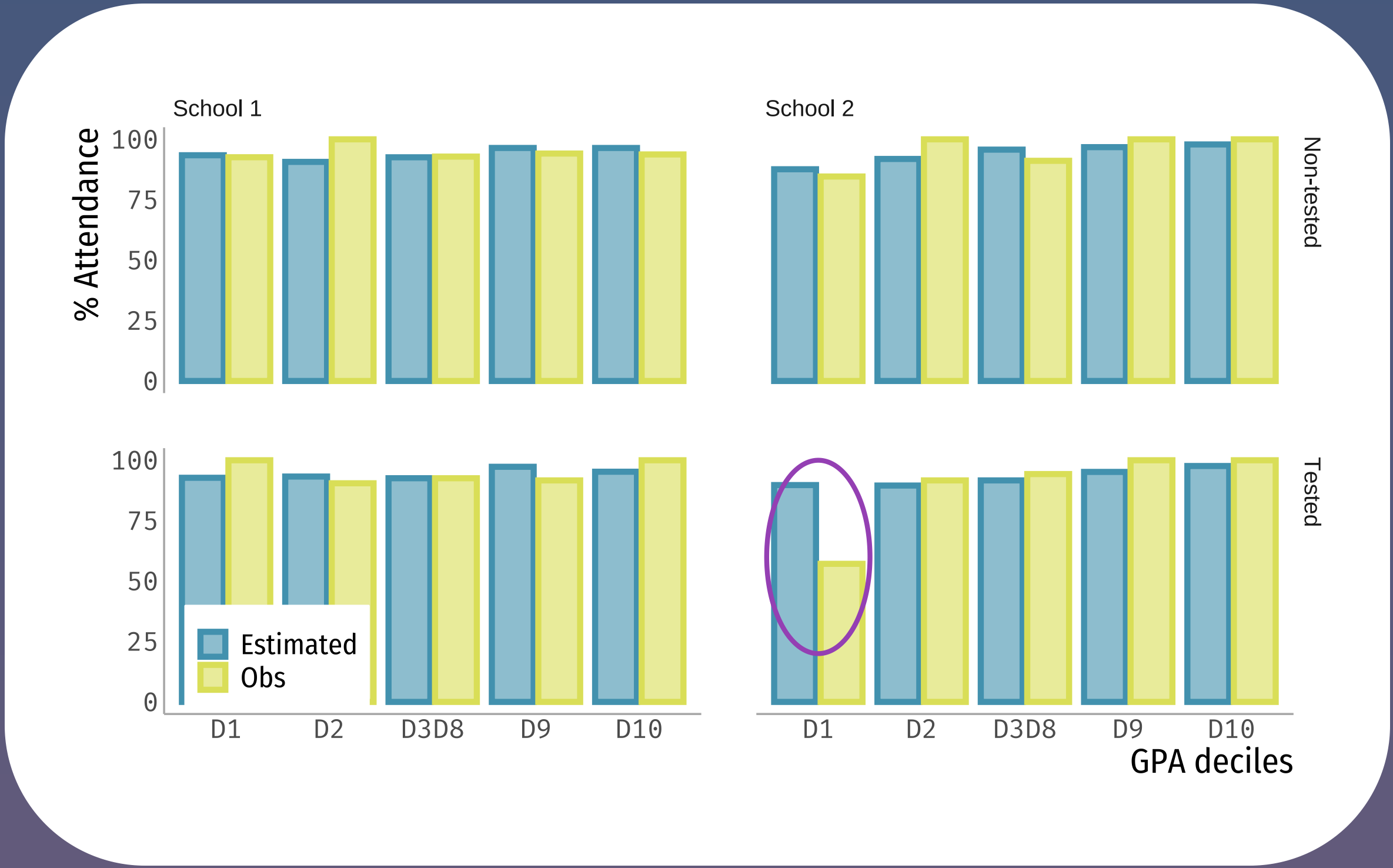
### Prediction of counterfactual attendance

- **XGBoost** for large student panel of daily attendance
  - Includes FE by day of the week, school, grade, and student. Also includes sibling's attendance (if any) and attendance lag.

Not only **low-performers attend less** on the day of the test in lower grades, but **high-performers attend more**



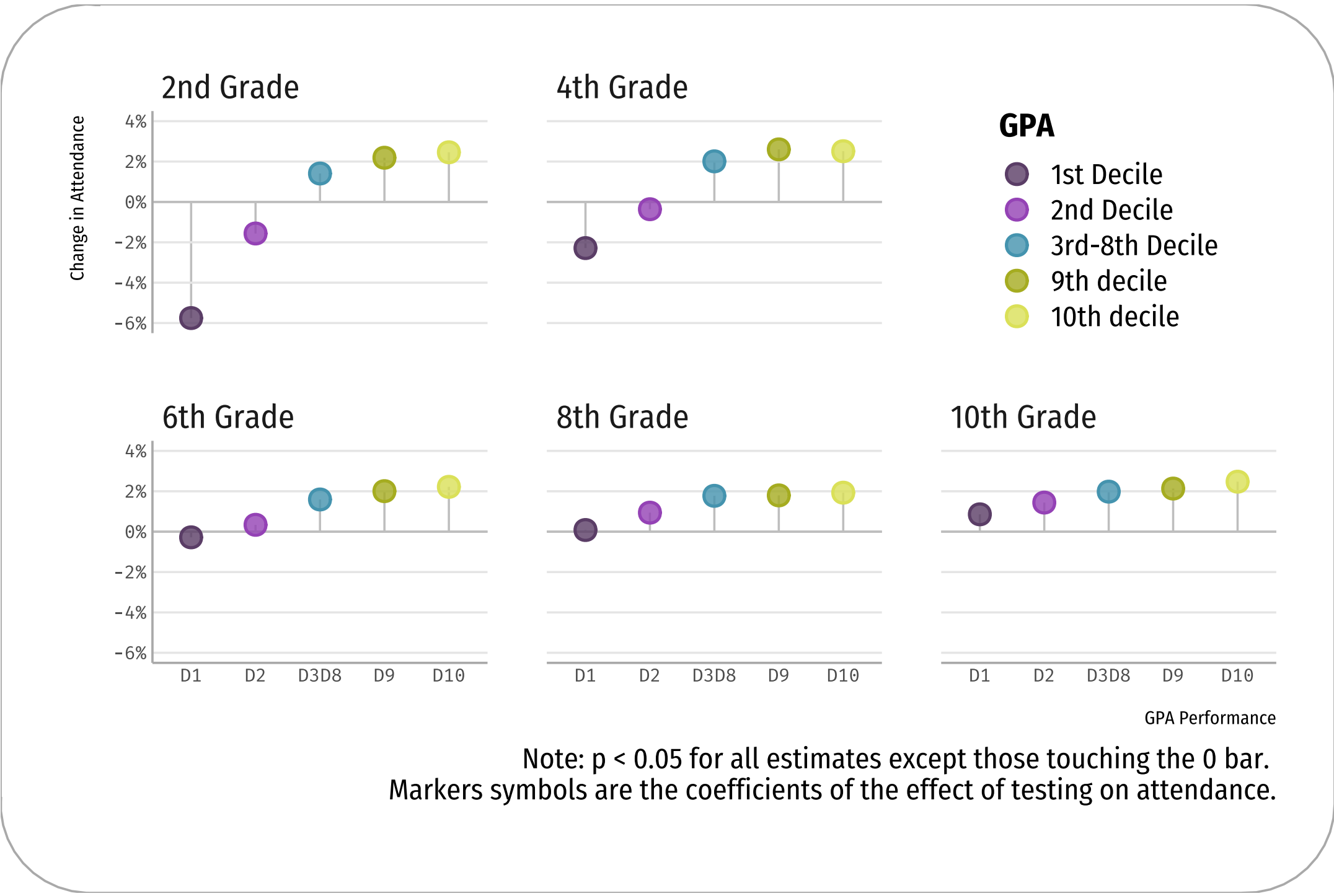
Using machine learning methods we can also **identify schools** more likely to **incentivize low-attendance** on bottom performers



- Identify types of schools by clustering on ( $Obs$  Attendance - Predicted Attendance) using **K-means**.
- Test different **imputation policies** and its consequences.

## Results

Students skip school on the day of the test. In lower grades, **lower-performers** attend less and **higher-performers** attend more, compared to a regular day. In higher grades, we **only observe action at the top of the distribution**



There is **important heterogeneity between schools**.

We use **K-means** analysis to identify **clusters of schools** according to their difference between predicted and observed attendance distribution. We find **two main clusters**, where one of them incentivizes the **exclusion of lower-performers**. Those schools are more vulnerable and have overall lower performance.

In terms of **imputation**:

- **Overall imputation** to match school population increases disparities
- **Imputation to match predicted distribution** is **inbetween** no imputation and imputation for all.

