Exploring teacher's repetition biases with a survey experiment

Ignacio Garijo-Campos

Antonio Alfonso

Pablo Brañas-Garza

Diego Jorrat

Ángel Sánchez

March 16, 2025

Abstract

Grade repetition has been widely debated due to its potential negative effects on students' self-esteem and long-term academic performance. While most research has focused on student outcomes, this study shifts the perspective to teachers' decision-making in grade repetition cases. We conduct a survey experiment in Spain, where the repetition rate is significantly higher than the OECD average. Teachers evaluate hypothetical student profiles and decide whether each student should repeat a grade. Our design allows us to analyze the impact of policy exposure, preference alignment, and cognitive biases on teachers' decisions. Additionally, we explore sociodemographic and behavioral factors associated with harsher repetition tendencies, including burnout levels, teaching experience, and institutional characteristics.

1 Introduction

Grade repetition has been regarded as a double punishment for students, who not only have to face educational failure, but also have to deal with being segregated from their colleagues. This can have a negative impact on the student's self esteem and perception of what they are capable of. In psychology, the term learned helplessness is often used to describe students that stop trying because they believe themselves to be uncapable of succeeding no matter the effort they put in. Indeed, literature points to little to no positive impact of grade repetition on students:

- Alet et al. (2013) show that repetition only has a positive impact on student test scores in the very short-run, while after 3 or 4 years the impact is significantly negative.
- Cabrera-Hernandez (2022) shows that eliminating repetition in Mexico led to a decrease of 30% in dropout rates without reducing the average student's performance. They show that not only grade repetition does not seem to be beneficial to the student, but that the *fear effect* is insignificant for other students.
- Finally, the OECD (2022) finds strong correlation between education systems with high repetition rate and low mathematical scores in standardized tests and other metrics.

While most studies on this phenomenon focus on students, we conduct an experiment on teachers, which allows to shed light on this issue from a novel perspective. The experiment is carried out in Spain, where 22% of students reported having experimented grade repetition, as opposed to the 9% OECD average (OECD, 2022).

First, we run a survey experiment where teachers are presented with different student profiles (according to 6 binary characteristics) and asked if the student should repeat grade or not. They repeat the task 8 times (8 students). We name this device as the *repetition task or 'cards game'*, and it allows us to estimate the harshness level of the teacher by comparing their decisions with those of the rest of the sample. This simple setting is presented for the control group.

In the treatment groups, teachers are randomly assigned to education systems (1 of 3 policies) before or after making the repetition task. They are also asked to rank the policies as *favorite*, *neither-favorite-nor-least-favorite* and *least favorite* (hereafter F/∼/LF).

Therefore, there are 3 elements to this experiment: The repetition task, the educational policy (3 scenarios) and the ranking over policies (F/∼/LF). Figure 1 summarizes the design described below:

Group I: Control Group. Teachers fill the repetition task and then select their preferences over the policies

Group II: Policy treatment. Teachers are randomly assigned to one of the 3 policies. Then, they are asked to complete the repetition task. Finally, they choose their preferences over the policies

Group III: Revelation Treatment. Teachers are asked to fill their preferences first, and then they are randomly assigned to a policy. Finally, they fill the cards with the repetition task.

Group IV: Awareness Treatment. This treatment is identical to the revelation one with an additional (randomly assigned) priming sentence: '*this is your F/∼/LF policy*'.

The experimental design allows us to make 3 main contributions. Firstly, we will be able to characterize *harder* teachers with both sociodemographical and behavioral characteristics. This is possible because teachers answer different questions before

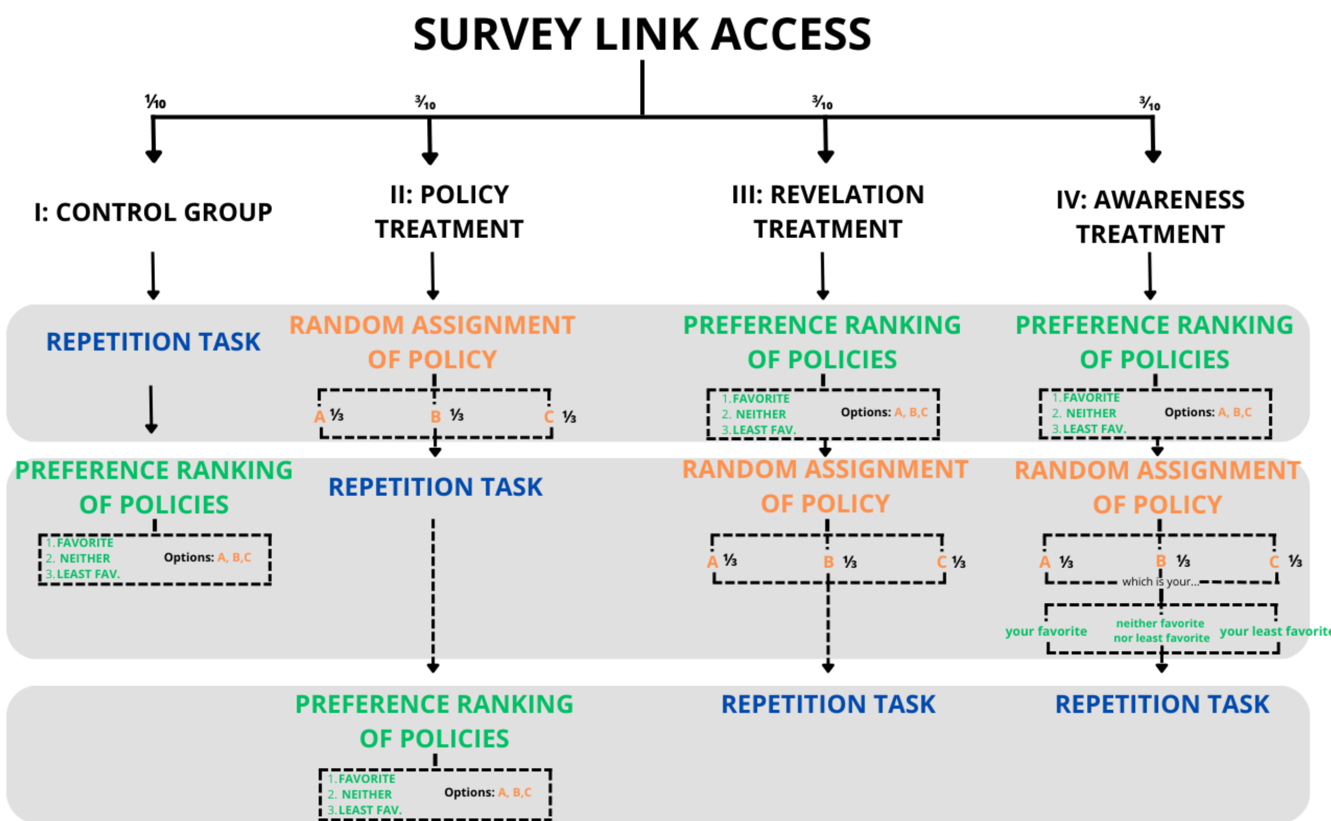

Figure 1: Experimental design

and after the experiment, and we will be able to tell their gender, age, experience, grade, area of expertise, ownership of the school in which they work, their burn-out level, confidence in the educational system, and confidence in meritocracy.

Secondly, we will be able to determine which student characteristics are more important for teachers when making repetition decisions. This is because the card game contains 6 characteristics of the student.

Finally, and most importantly, we will answer different questions (from now on, *studies*) regarding the teacher *harshness*, which is defined by comparing the teacher's decisions across treatments. These studies are connected to the different treatments, and allow us to explore important questions as if teachers are *harsher* when they are forced into an educational policy that does not align with their preferences.

2 Intervention & Experimental design

This is a one stage experiment, meaning that teachers are only contacted once. The intervention requires access through a link compatible for both mobile and desktop that will be sent by email to all teachers by the regional government.

It is worth noting that the teachers will be asked to answer two surveys, one before the experiment and one after its completion. The total estimated time of completion is of 18 minutes.

2.1 Surveys

The first survey (Appendix A) retrieves sociodemographical information as well as data of the type of school, course and classes they teach. They are also asked about labor information such as the years of experience or the type of contract they hold.

The second survey (Appendix B) positions the teacher behaviorally. While it asks many interesting questions like the teacher's confidence on meritocracy or their self-perceived level of empathy, arguably the most compelling question delves into the teacher's *burn-out* level (Q21).

This is especially interesting since it allows us to explore if the level of *burn-out* is correlated with repetition decisions. Moreover, we can investigate what causes teachers to be *burnt-out*.

Between these two surveys, the teacher goes through the randomized experiment as shown in figure 1. This part of the intervention is divided in different phases, which are presented in a different order depending on the treatment assigned to the teacher. The phases are:

1. **Preference Ranking:** the teacher faces the three policies and is asked

to rank them according to their preference through a vertical slider. This slider displays the numeric position in which the policy is being ranked (1 for favorite and 3 for least favorite). The original position of the three policies is random for all treatments, so as to avoid conditioning.

2. **Random assignment of policy:** teachers are assigned to one of the three policies randomly, with probability 1/3. Teachers in the control group do not go through this part, as they are not to be affected by policy assignment. The policies are displayed in appendix D.

3. **Repetition task/card game:** each card represents a hypothetical student with 6 binary characteristics with a fixed order. This means that there are 64 (2^6) possible hypothetical students. Teachers face 8 students each, meaning that there should be around $N/8$ observations of teachers' decisions on hypothetical student i. Those characteristics are reflected in Table 1.

Table 1: Card game scenario

	Option 1	Option 2
Gender	Male	Female
Migrant	Yes	No
Failed 3 or more subjects	Yes	No
Lacks mathematical and/or linguistic abilities	Yes	No
Has committed a serious or very serious infraction or has been suspended	Yes	No
Frequently misses class without justification	Yes	No

Before treatment assignment, the card game is explained to all subjects. Additionally, teachers are shown two examples: one of an ideal student with no negative characteristics and another of a hypothetical student exhibiting all the characteristics associated with grade repetition.

This approach is beneficial for two reasons. First, it ensures that the game is clearly presented and thoroughly explained to the subjects. Second, the results serve as an attention check to screen participants.

2.2 Sampling plan

This experiment will be sent via email to all primary and high schools of the region of Castilla La Mancha (Spain) by the 'Consejería de Educación, Cultura y Deportes' (Education, Culture and Sports Regional Authority).

The schools and high schools are then responsible for resending the link to their teachers. Every teacher with internet access is then qualified to answer the survey, which can be answered both by phone or desktop.

While we do not have a current exact number of teachers of the region, numbers from the 2022-2023 school year are available at the regional government's website (Consejería de Educación, 2024). During that school year, there were 35 thousand teachers in the region, from which 26 thousand taught primary or secondary levels, meaning they were qualified for the survey. 76.12% of those 26 thousand taught at public schools, and 67.46% were women. Even though we are unaware of the number of teachers today, we can expect a similar number.

Based on our power analysis to detect an effect at $p < 0.05$ level with 80% power, we have estimated the following MDEs (in SD) depending on the number of teachers that complete the survey (table 2):

Table 2: Tabla de MDE en función de n

n	**MDE (in SD)**
2500	0.24
3500	0.20
4500	0.18
5500	0.16
6500	0.15
7500	0.14

These calculations were made with $\alpha = 0.05$, power $= 0.8$, $R^2 = 0.1$, 10 experimental arms, and $SD = 0.10$

3 Studies

This section details the hypotheses of the experiment, articulated in the form of null hypotheses in the sense that they are all expressed in terms of no impact.

Here, treatment assignment will refer to the treatment in which the teacher has been allocated (control, policy, revelation, awareness); while policy assignment refers to the 3 possible allocations of policy (A, B and C).

Study I: Control vs Policy Treatment

The Control Group faces the repetition task and then rank the policies; while in the Policy Treatment, teachers only go through the repetition task and policy ranking after being randomly assigned to a policy (Figure 1).

This design allows us to test the aforementioned hypotheses:

$H_{1|1}$: **Teachers' harshness is not correlated with being assigned to policies.** This means that teachers assigned to the Policy Treatment are not harder than those from the Control Group.

$H_{1|2}$: **Teachers' harshness is not correlated with being assigned to a specific policy.** This means that the interaction of treatment assignment and policy assignment has no impact on teacher's hardness.

$H_{1|3}$: **Teachers' harshness is not correlated with the alignment of teacher's preferences and their policy assignment.** This means that teacher's assigned to their favorite nor least favorite policies are neither harder nor softer.

Study II: Revelation vs Policy Treatment

The difference between the Revelation and Policy Treatment is that the latter ranks the policies before being randomly assigned to one. This means that, by the time they are assigned to a policy, they have already decided their preferences.

Here, our hypotheses are:

$H_{2|1}$: **Teachers' harshness is not correlated with being asked to rank the policies before policy assignment.** This means that teachers assigned to the Revelation Treatment are not harder than those from the Policy Treatment.

$H_{2|2}$: **Teachers' harshness is not correlated with being assigned to a specific policy after having ranked them.** This means that the interaction of treatment assignment and policy assignment (involvement) has no impact on teacher's hardness.

$H_{2|3}$: **Teachers' harshness is not correlated with the alignment of teacher's preferences and their policy assignment when teacher's have ranked the policies before assignment.** This means that teacher's assigned to their favorite or least favorite policies after having ranked them are neither harder nor softer that those who had not revealed them before the ranking.

Study III: Endogenenous vs Awareness Treatment

Here, the difference between the Awareness and the Revelation Treatment is that the latter is explicitly told if the policy they are being assigned to is their 'favorite'/'least favorite'/'neither favorite nor least favorite'.

This leads us to:

$H_{3|1}$: **Teachers' harshness is not correlated with acknowledging their opinion of the policy they are being assigned to.** This means that teachers assigned to the Awareness treatment are not harder than those from the Revelation one.

$H_{3|2}$: **Teachers' harshness is not correlated with being made aware of their opinions when assigning a specific policy.** This means that the interaction of treatment assignment and policy allocation has no impact on teacher's hardness.

$H_{3|3}$: **Teachers' harshness is not correlated with the alignment of teacher's preferences and their policy assignment in a setting where they are aware of the alignment or disalignment.** This means that teachers are neither harder nor softer when assigned to their favorite or least favorite policies with full awareness of their opinion on that policy.

Study IV: Additional analyses

This experimental design also allows us to explore some additional hypotheses which do not focus on the level of harshness of the teacher as an outcome variable.

$H_{4|1}$: **Current decisions are independent of previous decisions in the repetition task.** This is a test on *moral licensing*, which will tell us if teacher's compensate their previous decisions with the present ones.

$H_{4|2}$: **Policy preferences are orthogonal to policy assignment.** This is a test on *conformity*. Here, we explore if being assigned to a policy has any impact in the later revealed preferences for treatment II.

$H_{4|3}$: **Treatment assignment has no effect on policy preferences.** This tests if the treatments distorted teachers' policy preferences.

For the sake of clarity, table 3 shows which hypotheses are tested with which samples:

Table 3: Hypotheses by treatment

Hypothesis	Control	Policy	Revelation	Awareness
$H_{1\|1}$	X	X		
$H_{1\|2}$	X	X		
$H_{1\|3}$	X	X		
$H_{2\|1}$		X	X	
$H_{2\|2}$		X	X	
$H_{2\|3}$		X	X	
$H_{3\|1}$			X	X
$H_{3\|2}$			X	X
$H_{3\|3}$			X	X
$H_{4\|1}$	X	X	X	X
$H_{4\|2}$		X		
$H_{4\|3}$	X	X	X	X

4 Analysis plan

4.1 Variable of interest: teacher's *harshness*

Teacher's *harshness* can be measured in several ways. We have gathered a number of options:

Number of repetitions

Since they are all shown the same amount of cards, with characteristics that are randomly assigned, a high number of repetitions is an intuitive way of identifying *harsher* teachers,

$$H_j^a = \sum_{i=1}^{8} R_{i,j} \tag{1}$$

Where $H_j^a \in [0, 8]$ is the sum of the binary decisions (repetition$= 1$, non repetition$=$ 0) of teacher j for the student cards i from 1 to 8.

Average deviation

The same card is shown to several different teachers ($N/8$ as discussed before). Then, if we identify one hypothetical character i as a specific combination of characteristics, one could estimate equation 2:

$$H_j^b = \frac{\sum_{i=1}^{8} \left(R_{i,j} - \bar{R}_{i,-j} \right)}{8} \tag{2}$$

Where $R_{i,j}$ is a binary variable with the repetition decision of teacher j for card i; and $\bar{R}_{i,-j} \in [0, 1]$ is the rate of repetition for that card i for all teachers that were given that same card, excluding teacher j.

For instance, if 60% of teachers decide to make student i repeat course, and teacher

j does not, $\left(R_{i,j} - \bar{R}_{i,-j}\right) = (0 - 0.6) = -0.6$. The average of the 8 results for professor j indicates their level of harshness. Consequently, $H_j^b \in [-1, 1]$

This measure is more context based since it compares the decisions of the teacher with the average decision of other teachers with the same card. It is thus a relative measure of *harshness*.

Deviation with respect to predicted probability

Since all cards include the same 6 characteristics, we can estimate a logit model (equation 3) that predicts the probability of repetition of a specific card according to their characteristics. We also control for the level of education taught by the teacher (primary/secondary) since regulations are different for each level, and for the ownership of the school (private/public/charter), as this has a great impact on repetition patterns.

$$\log\left(\frac{P_i}{1 - P_i}\right) = \sum_{k=1}^{6}(\delta_k M_k) + \mu L + \upsilon O \tag{3}$$

Where δ_k is the coefficient for characteristic M_k; μ is the coefficient for the level of education taught by the teacher L; and υ is the coefficient for the variable that accounts for the ownership of the school O. The predictions of this model are then used in a similar way as the rate of repetition of a specific card in equation 2.

In this case, we estimate:

$$H_j^c = \frac{\sum_{i=1}^{8}\left(R_{i,j} - \hat{R}_i\right)}{8} \tag{4}$$

Where $\hat{R}_i$ is the predicted probability of repetition of card i according to the model estimated by equation 3. Similarly to H_j^b, $H_j^c \in [-1, 1]$

4.2 Characterizing the teacher

One of the main contributions of this study does not require causality to be valuable. Indeed, being able to characterize teachers with higher repetition propensity is extremely relevant for the policy-maker.

Here, we will use random forest regressions to characterize hard teachers using the numeric continuous measures H_j^b and H_J^c as outcome variable.

This methodology allows us to estimate feature importance, i.e., we can analyze which variables are better at explaining higher levels of harshness. The variables that will be used for this are those of appendix C. We will perform this analysis with those variables included in $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ separately. This is because there might be some missing values in the latter as those questions are answered in the end.

The final manuscript will include the plotting the level of harshness against the most important variables so the correlation can be observed graphically.

4.3 Empirical strategy: studies

We will estimate regression equations for each study, using H_j^b and H_j^c ($\boldsymbol{H_j}$ from now on) as dependent variables. We will also estimate them with and without covariates (see X_1 in appendix C for further detail) to check if these variables could net out the possible significance of the treatment coefficients. For the sake of simplicity, the following equations will not include the covariates and their coefficients, even though results will be shown with and without them.

Moreover, we cluster standard errors at the school level. Due to personal data protection law, we cannot directly identify the school where the teacher works at, but we can infer it considering we have a large amount of information on the school from Questionnaire 1 (Appendix A). Variables like school's postal code, the type of ownership, the levels of education taught at the school, and the principal's gender should allow us to construct an anonymized school id. Since we do not know *a priori* the number of respondents, if the number of clusters is too small (lower than 25), we will compute standard errors using bootstrap.

Study I: Policy treatment vs. Control

This study examines $H_{1|1}$, $H_{1|2}$, and $H_{1|3}$.

$\underline{H_{1|1}:}$

$H_{1|1}$ states that teachers harshness level should not depend on being exposed to policies. Then, we can estimate equation 5:

$$\boldsymbol{H_j} = \lambda_{1|1} D_1 \tag{5}$$

Where $\lambda_{1|1}$ is the coefficient for policy treatment binary variable D_1.

Discussion: A negative $\lambda_{1|1}$ would mean that being exposed to the policies has a negative effect on the harshness level of the teachers. This could be the case if teachers are not happy with being *forced into* a policy, no matter whether they like it or not. This possible outcome will also be clarified with $H_{4|2}$, which explores if being assigned to a policy has an impact of the probability of liking/disliking it.

$\underline{H_{1|2}}$

$H_{1|2}$ differs from $H_{1|1}$ in the sense that it does not explore the possible treatment effect of the assignation of treatment, but rather the possible treatment effect of policy assignment.

For $H_{1|2}$, we can answer by estimating equation 6:

$$\boldsymbol{H_j} = \boldsymbol{\lambda_{1|2}^{\Omega'} D_{\Omega'}} + \boldsymbol{\gamma'_{1|2} F'} \tag{6}$$

Where $\boldsymbol{\lambda_{\Omega'}}$ is a vector of coefficients for the vector of binary variables created by the categorical variable $\boldsymbol{D_{\Omega'}} \in [0, 1, 2, 3]$ depending on if the subject was assigned to the Control Group and thus no policy ($D_{\Omega'} = 0$), Policy A ($D_{\Omega'} = 1$), Policy B ($D_{\Omega'} = 2$), or C ($D_{\Omega'} = 3$). In addition, vector $\boldsymbol{\gamma' F'}$ controls for the favorite policy of the teacher.

Discussion: significant coefficients would imply that teachers act differently depending on which policy they receive from the assignment, regardless if they like it or not. Insignificant coefficients would point in the direction that teachers' decisions are not explained by the policy under which they are asked to act. Other factors could be personal preferences, sociodemographical characteristics, working environment (school), etc.

$\underline{H_{1|3}}$

For $H_{1|3}$, we study if the alignment of teacher's preferences and policy assignment has an impact on their harshness level.

For this, we restrict the sample by eliminating those who were allocated to a *neither-favorite-nor-least-favorite* policy. That being said, we can estimate equation 7:

$$\boldsymbol{H_j = \phi P + \lambda^{\Omega}_{1|3} D_{\Omega}} \tag{7}$$

Where $\boldsymbol{\phi}$ is a vector of coefficients for the binary variables created from $\boldsymbol{P} \in [0, 1, 2]$ if teacher j was assigned to control group ($P = 0$), their favorite policy ($P = 1$), or their least favorite policy ($P = 2$). $\boldsymbol{\lambda^{\Omega}_{1|3}}$ is a vector of coefficients for the binary variables created by the categorical variable D_{Ω} which can take value 1 (Policy A), 2 (B) or 3 (C).

Discussion: This study will help us clarify if teachers are harder under a situation they are not keen on, or when they are forced into a policy they do not appreciate. This will be discussed further in studies II and III, where teacher's have seen and ranked the policies before being assigned to them, meaning they have formed and revealed their preferences by the time they receive a policy.

Study II: Revelation treatment vs. Policy treatment

This study examines $H_{2|1}$, $H_{2|2}$, and $H_{2|3}$, comparing treatments II and III.

In treatment II, teachers are randomly assigned a policy, but they only get to see and rank the other two once the card game is over. In treatment III, however, they are shown and asked to rank the 3 policies before the policy assignment and the repetition task.

This means that treatment III teachers already know if a policy is their favorite or least favorite when they receive it.

$\underline{H_{2|1}:}$

$H_{2|1}$ affirms that teacher's harshness is not correlated with having already ranked the 3 policies before being designated to one.

Since this hypothesis is similar to $H_{1|1}$ but comparing different treatments, the estimation strategy is also similar (eq. 8):

$$\boldsymbol{H_j} = \lambda_{2|1} D_2 \tag{8}$$

Where $\lambda_{2|2}$ is the coefficient for treatment variable D_2, which takes value 1 when the teacher belongs to treatment III and 0 when they belong to treatment II.

Discussion: a positive significant coefficient would mean that a teacher in treatment III is harder than a teacher in treatment II. This could be the case if the teachers that are unhappy with the policy they have been assigned to in treatment III boycott the repetition game. This is because teachers in treatment II have not yet formed their opinion on the policies when they do the repetition task.

$\underline{H_{2|2}}$

$H_{2|2}$ states that the policy a teacher is assigned to does not have an impact on their harshness when comparing treatments II and III.

This hypothesis is also similar to $H_{1|2}$ but in this case the 'control group' is also assigned to a policy, so the estimation of equation 9 is slightly different:

$$\boldsymbol{H_j} = \lambda_{2|2} D_2 + \boldsymbol{\lambda^{\Omega}_{2|2} D_{\Omega}} + \boldsymbol{\lambda_{2|2}} D_2 \boldsymbol{D_{\Omega}} + \boldsymbol{\gamma'_{2|2} F'} \tag{9}$$

Here, the policy assignment variable ($\boldsymbol{D_{\Omega}}$) and the treatment assignment variable (D_2) interact with each other to study the interaction between policy and treatment random assignment.

Discussion: significant $\lambda_{2|2}$ coefficients could point in the direction that some policies have an effect on the harshness when presented after preferences are revealed. Since we control for policy preferences ($\gamma'_{2|2} F'$), this could not be an explanation of a possible significant coefficient. Indeed, coefficients significantly different than zero would suggest that that there is retaliation (positive compensation) when teacher's are assigned a policy they have previosly shown to dislike (like). This commitment mechanism is further explored in $H_{2|3}$, where we focus in the alignment of policy preferences and assignment.

$\underline{H_{2|3}}$

$H_{2|3}$ states that the alignment between teacher's preferences and their policy assignment does not impact the level of harshness when preferences are revealed prior to policy assignment.

To explore $H_{2|3}$, we run 2 regressions per dependent variable: one with the teachers from treatments II and III that were assigned their favorite policy ($n_F^{2,3}$); and one with the teachers from treatments II and III that were assigned their least favorite policy ($n_{LF}^{2,3}$). Equation 10 describes our estimation strategy:

$$\boldsymbol{H_j} = \lambda_{2|3} D_2 + \boldsymbol{\gamma'_{2|3} F'} \tag{10}$$

where $\lambda_{2|3}$ symbolizes the coefficient for binary variable D_2 that takes value 1 when the subject belongs to treatment III and 0 when they belong to treatment II.

Discussion: a positive significant coefficient $\lambda_{2|3}$ for subsample $n_F^{2,3}$ would suggest that teachers are less harsh when given what they have chosen as their favorite policy. Similarly, a negative significant coefficient $\lambda_{2|3}$ for subsample $n_{LF}^{2,3}$ would suggest that teachers retaliate when given a policy they have stated as their least favorite. If proven, we would call that *boycott effect*.

Study III: Revelation treatment vs. Awareness treatment

Study III explores hypotheses $H_{3|1}$, $H_{3|2}$, and $H_{3|3}$ by comparing treatments III and IV.

Treatment IV differs from III in a very subtle manner: treatment IV teachers are explicitly told if the policy they are being assigned to is their favorite, least favorite, or neither favorite nor least favorite. For instance, a teacher who is assigned to their least favorite policy will see the message ‘You are informed that the implemented policy is your third option’ if they belong to treatment IV.

$\underline{H_{3|1}}$

Just like in $H_{1|1}$ and $H_{2|1}$, $H_{3|1}$ explores if there are treatment effects in the treatment assignment, i.e. if teachers in the Awareness Treatment (IV) are harder than those in the Revelation Treatment (III).

The equation 11 is thus:

$$\boldsymbol{H_j} = \lambda_{3|1} D_3 \tag{11}$$

Discussion: a coefficient $\lambda_{3|1}$ significantly different than zero would convey treatment effects. This could be true, for instance, if those in the Awareness Treatment who are given their least favorite policy retaliate much more than those in the Revelation Treatment. This is explored further in the following hypotheses.

<u>$H_{3|2}$</u>

This hypothesis states that there are no treatment effects in the interaction between treatment assignment and policy assignment. Meaning that those that are assigned to, for instance, policy A in the Awareness Treatment, are not harsher nor more lenient than those assigned policy A in the the Revelation Treatment. Equation 12 details the regression that will be used to determine this:

$$\boldsymbol{H_j} = \lambda_{3|2} D_3 + \boldsymbol{\lambda^{\Omega}_{3|2} D_{\Omega}} + \boldsymbol{\lambda_{3|2}} D_3 \boldsymbol{D_{\Omega}} + \boldsymbol{\gamma'_{3|2} F'} \tag{12}$$

Discussion: Coefficients $\lambda_{3|2}$ significantly different than zero could be possible if, for instance, the level of implication of a teacher with the policy assigned to them varies between Treatments III and IV. It could also be the case if the retaliation of those assigned to their least favorite policy is greater in the Awareness treatment.

<u>$H_{3|3}$</u>

Here, we will check $H_{3|3}$, to see if lenience (boycotting) is increased when professors are told directly if they are being assigned their (least) favorite policy. To estimate the effect, we use a similar strategy as in study II, meaning we carry out two different regressions for each outcome variable. In this case, we use the people in the Awareness Treatment as treatment and the ones from the Revelation Treatment as control. The estimation is described in equation 13.

$$\boldsymbol{H_j} = \lambda^{F,LF}_{3|3} D_3 + \boldsymbol{\gamma'_{3|3} F'} \tag{13}$$

Discussion: Here, a coefficient $\lambda^{F}_{3|3}$ ($\lambda^{LF}_{3|3}$) significantly different than zero would mean that teachers are significantly more lenient (harsh) when assigned to a policy they like (dislike) with an explicit message for it. This would have relevant implications for the policy maker as it would imply that people (in this case, teachers) act differently when they feel heard (ignored).

Study IV: Additional Analyses

$\underline{H_{4|1}:}$

To study $H_{4|1}$ we will estimate a logit AR1 model which will explore if the probability of passing/failing can be predicted by the previous decisions of the same teacher, i.e., if they try to compensate past actions, as in Brañas-Garza et al. (2013). Equation 14 details our identification strategy:

$$\gamma_{j,i,t} = \rho_{t-1}\gamma_{j,i,t-1} + \rho_{t-2}\gamma_{j,i,t-2} \tag{14}$$

Where $\gamma_{j,i,t} = 1$ if teacher j fails student i at interaction $t \in [1, 8]$, and 0 if the decision is passing the student.

$\underline{H_{4|2}:}$

To study $H_{4|2}$ we examine if teachers are more prone to disliking the policy they are being assigned to. Using the subsample of subjects from the policy treatment,

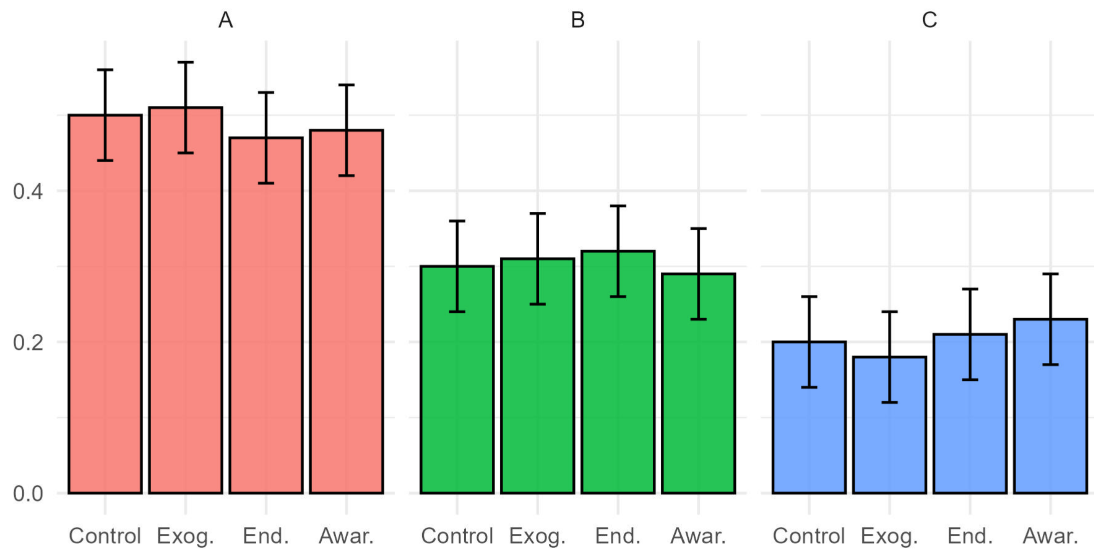

Figure 2: $H_{4|3}$ simulation

we can estimate this with defining three logit models (one for each policy) such as equation :

$$\log\left(\frac{P_i^{\Omega}}{1-P_i^{\Omega}}\right) = \lambda_{4|2} D_{\omega} \tag{15}$$

Where the outcome variable is the probability of choosing a policy Ω as favorite; $\lambda_{4|2}$ is the coefficient for binary variable D_{ω}, that takes value 1 when the subject has been assigned to the same policy Ω and 0 when they have been assigned to a different one.

$H_{4|3}$:

Finally, for $H_{4|3}$, we study if preferences are dependent on the treatment. To study this, we will show a figure with the probability of assigning a policy Ω as favorite or least favorite by treatment with their standard errors, as in hypothetical figure 2.

4.4 Inclusion and exclusion criteria

We will exclude the top and bottom 1.75% in terms of time taken to complete the survey. This is because we want to avoid answers from participants who answered mindlessly (too fast), and participants that were not paying attention (too slow).

Participants are allowed to exit the experiment and continue even days later without having to start again. Once they answer the last question, it will be not possible for them to answer it again from the same device and browser.

As abovementioned, before treatment assignment, teachers are explained the repetition task and presented with two examples (one of an ideal student and one that meets all the usual characteristics of a student that repeats grade). This works as an attention test, and thus we will eliminate observations that fail the ideal student, since there are no reasons for the subject to make this hypothetical student repeat. Nonetheless, we will not eliminate observations of those teachers who do not fail the non-ideal student. This is because some teachers may not agree with grade repetition at all, which does not convey irrationality or lack of understanding of the task.

On the other hand, we include in our analysis all teachers that finish the repetition task, and thus the first survey. Therefore, those who do not complete the second survey will still be considered in our analysis. Additionally, section 4.2 estimates random forest regressions to characterize harder teachers. This will be performed twice: one with the covariates contained in X_1 and one with the covariates in X_2, both listed in appendix C. The vector X_2 includes answers to questions that belong to the second survey, meaning that for that part, only the subset of the sample that answered all these questions will be used for the estimation.

A Questionnaire 1

1. I am a teacher of:

   - **Primary Education**

     ☐ Kindergarten

     ☐ 1st Grade

     ☐ 2nd Grade

     ☐ 3rd Grade

     ☐ 4th Grade

     ☐ 5th Grade

     ☐ 6th Grade

     ☐ Other: ______________

   - **Secondary Education**

     ☐ 7th Grade (1º ESO)

     ☐ 8th Grade (2º ESO)

     ☐ 9th Grade (3º ESO)

     ☐ 10th Grade (4º ESO)

     ☐ High School (Bachillerato)

     ☐ Vocational Training

     ☐ Other (specify): ______________

   - **Subjects taught:**

☐ Biology and Geology

☐ Classics

☐ Art

☐ Economics

☐ Physical Education

☐ Philosophy

☐ Physics and Chemistry

☐ French

☐ History and Geography

☐ English

☐ Language and Literature

☐ Mathematics

☐ Music

☐ Counseling

☐ Technology

2. I have dedicated _____ years to teaching.

3. I have taught at this school for _____ years.

4. Previously, I have worked at _____ educational institutions.

5. I identify as:

   ☐ Male

- ☐ Female
- ☐ Non-binary
- ☐ Prefer not to answer

6. I am _____ years old.

7. Last academic year, I was a homeroom teacher for a group:

   - ☐ Yes
     - The grade I tutored was: _______________
     - That group had _____ students.
     - From that group, how many students did you think might repeat the grade? _____
     - How many actually repeated the grade? _____
   - ☐ No

8. What is your school's postal code? _______________

9. Your school's type of ownership:

   - ☐ Public
   - ☐ Charter
   - ☐ Private

10. Your employment status:

    - ☐ Permanent Teacher (public servant)
    - ☐ Temporary Teacher (public servant)

- ☐ Intern
- ☐ Contracted (Indefinite)
- ☐ Contracted (Temporary)
- ☐ Internship Contract
- ☐ Other: _______________

11. What levels of education are taught at your school? (Select all that apply)
    - ☐ Preschool
    - ☐ Elementary
      - How many 6th grade classes? _____
    - ☐ Secondary (Middle School)
      - How many 7th grade classes? _____
    - ☐ High School
    - ☐ Vocational Training (Basic Level)
    - ☐ Vocational Training (Intermediate Level)
    - ☐ Vocational Training (Higher Level)
12. How many class groups do you have? _______________
13. How many hours did you substitute for a colleague last week? _______________
14. How many students are in your largest class? _______________
15. How many absentee students do you have in total? _______________
16. How many disruptive students do you have in total? _______________

17. Does your school have specific programs for vulnerable students? (e.g., Prepara-T, Titula-S+, PISE+, etc.)

    ☐ Yes

    ☐ No

18. The school principal is:

    ☐ Male

    ☐ Female

    ☐ Non-binary

    ☐ I don't know

19. How long has the principal been in their position?

    ☐ Less than 1 year

    ☐ 1 to 4 years

    ☐ 5 to 8 years

    ☐ More than 8 years

- Open question: What do you think of eliminating the certificate of completion of secondary education and replace it with a report on the student skills.

- Open question: What do you think of legally banning grade repetition?

B Questionnaire 2

21. Rate your level of agreement with the following statements on a scale from 0 (strongly disagree) to 10 (strongly agree):

	0	1	2	3	4	5	6	7	8	9	10
My work has a positive impact on students.											
Some students pass the year without achieving the necessary competencies.											
Students who pass the year are adequately prepared for the next academic level.											
Too many resources are allocated to repeating students.											
No matter what we do, additional resources for repeating students are ineffective.											

22. If you had to distribute the responsibility for grade repetition among students, families, faculty, and the education system, what percentage would you assign to each? *(The total must not exceed 100%.)*

	%
Students	
Families	
Faculty	
Education System	
Total	

23. To what extent does the following statement describe you? *"I am an empathetic person."* *(Rate using a scale from 1 = Not true for me to 5 = Very true for me.)*

24. Some people believe that an individual's economic status depends almost entirely on family background, social connections, or luck rather than effort, education, and professional merit. Others think that effort, education, and merit are what truly matter. On a scale from 1 to 10, where 1 represents luck (family background and connections) and 10 represents effort (education and merit), how much do you think effort influences a person's economic status in Spain?

Luck 1 2 3 4 5 6 7 8 9 10 **Effort**

25. If your faculty consisted of 10 teachers (excluding yourself), how many do you think would agree with the following?

   - ☐ Banning grade repetition: __________
   - ☐ Replacing the high school diploma with a competency report for each student: __________

26. Rate your level of agreement with the following statements on a scale from 1 (strongly disagree) to 6 (strongly agree):

	1	2	3	4	5	6
There are many social norms in our society that people are expected to follow.						
People in our society always know what is expected of them in different situations.						
In our society, people generally agree on what behaviors are appropriate or inappropriate in most situations.						
People in our society have a lot of freedom to decide how to behave.						
In our society, if someone acts inappropriately, others quickly disapprove.						
People in our society almost always follow social norms.						

27. How has the grade repetition rate at your school changed compared to 15 years ago?

    ☐ Higher

    ☐ Same

    ☐ Lower

C Covariates

Adding controls to a regression with randomized treatment assignation can increase the power of the estimation. Table 4 shows the lists of covariates we will include in our analyses:

Variable	**Potential Values**	$\boldsymbol{X}_1$	$\boldsymbol{X}_2$
Level of education	Primary, Secondary	Yes	Yes
Years in teaching	$(0, \infty)$	Yes	Yes
Years in the current school	$(0, \infty)$	Yes	Yes
Gender	Male, Female, Non-binary, Prefer not to answer	Yes	Yes
Age	$(18, \infty)$	Yes	Yes
School ownership	Public, Semi-private, Private	Yes	Yes
Employment status	Temporary (substitutes, trainees, and temporary contracts), Permanent (tenured teachers with permanent or provisional placement, indefinite employment contract)	Yes	Yes
Number of groups taught	$(0, \infty)$	Yes	Yes
Self-perception of positive impact on students	0-10	No	Yes
Self-perception of empathy	1-5	No	Yes
Belief in the importance of effort in economic success	1-10	No	Yes

Table 4: Description of variables and their possible values

D Policies

The following policies are the ones that will be randomly assigned to the teachers, from which they will have to choose which is their *favorite* and which is their *least-favorite*. The remaining one will be considered *neither favorite nor least-favorite*:

A: The student will receive educational reinforcement in a small group next school year (in a small group, with a support teacher, inside or outside the classroom).

B: The promotion criteria will be modified in the Evaluation Board: unanimity for grade retention, secret and anonymous voting.

C: All school staff will receive tailored training and teaching support in inclusive teaching methodologies and multi-level classroom management.

References

Alet, E., Bonnal, L., & Favard, P. (2013). Repetition: Medicine for a short-run remission. *Annals of Economics and Statistics*, 227–250.

Brañas-Garza, P., Bucheli, M., Espinosa, M. P., & García-Muñoz, T. (2013). Moral cleansing and moral licenses: Experimental evidence. *Economics & Philosophy*, *29*(2), 199–212.

Cabrera-Hernandez, F. (2022). Leave them kids alone! the effects of abolishing grade repetition: Evidence from a nationwide reform. *Education Economics*, *30*(4), 339–355.

Consejería de Educación, C. y. D. (2024). Profesorado [Accessed on March 14th, 2025]. https://www.educa.jccm.es/es/consejeria-educacion-cultura-deportes/estadistica-educativa/estadistica-2022-2023/2-profesorado

OECD. (2022). *Pisa 2022 results (volume ii): Learning during and from disruption.*