Training and Bureaucratic Performance: Evidence from Peruvian Judges

Chen, D.; Ramos-Maqueda, M.; Silveira, B.

Blavatnik School of Government
University of Oxford &
Development Impact Evaluation (DIME)
The World Bank

May 15, 2023

Table of Contents

Table of Contents

Motivation

- The training of public officials is one of the key dimensions governments use to improve bureaucratic performance
- For example., in 2017 alone, the U.S. allocated approximately 4% of its annual budget for personnel compensation and benefits, or around $10 billion, towards training civil servants
- Despite its significance, there is limited empirical research on effective methods to improve the training of public officials
- Particularly relevant in the judiciary, as slow and unreliable justice systems represent a key barrier to economic growth

Motivation

- We partner with the Judicial Academy of Peru. The Judicial Academy Training (JAT) program trains active judges and prosecutors seeking promotion every year
- We ask: Can providing online feedback to trainers enhance the quality of civil servant training and lead to improved public services for citizens?
- We implement an RCT for the 8 months of the JAT Program, covering 8 rounds of 22 classes with 604 judges and prosecutors

Results preview

- The intervention increased student satisfaction with the trainer and with the course by 0.1 standard deviations (SD)
- Treated students increased their grades in each course by 0.12 SD relative to the control group
- Judges who were treated increased their judicial efficiency [based on preliminary results]
  - We find an increase in the case clearance rate of judges, i.e., the ratio of cases resolved based on the number of cases filed
  - We find a reduction in the time to disposition of judicial cases

Results preview

The positive effects of the interventions are driven by female judges and prosecutors:

- When subsetting the data by gender, the effects are driven by female students
- We also find a reduction in traditional gender stereotypes, measured through Implicit Association Tests (IAT)
- The reduction in gender stereotypes is driven by male students in the program

Contributions to the literature

1. **On state effectiveness:** We demonstrate that methods that improve the quality of teaching to bureaucrats can have downstream effects on the delivery of public services, improving the efficiency of a slow justice system in our study (Banerjee et.al, 2021;Mehmood et.al, Baye and Wright, 2011)
2. **On the malleability of implicit attitudes:** We show that changing the way of teaching can impact implicit bias in high-stakes decision-makers, judges (Carlana, 2019; Alan et.al, 2018; Mehmood et al., 2021; Alan et al, 2020; Jayachandran, 2021)
3. **On virtual learning:** expand the context of trainer feedback interventions to the virtual context and high-stakes decision-makers (Kirabo, 2012; Rockoff, 2004; Allen et al, 2011)

Table of Contents

Context of the study

- We partnered with Peru's Judicial Academy Training (JAT) program.
- Between May and December 2020, the JAT trained 604 active judges and prosecutors across 22 classes in virtual sessions
- AMAG randomized whether trainers receive an online assessment and feedback program

Overall Timeline

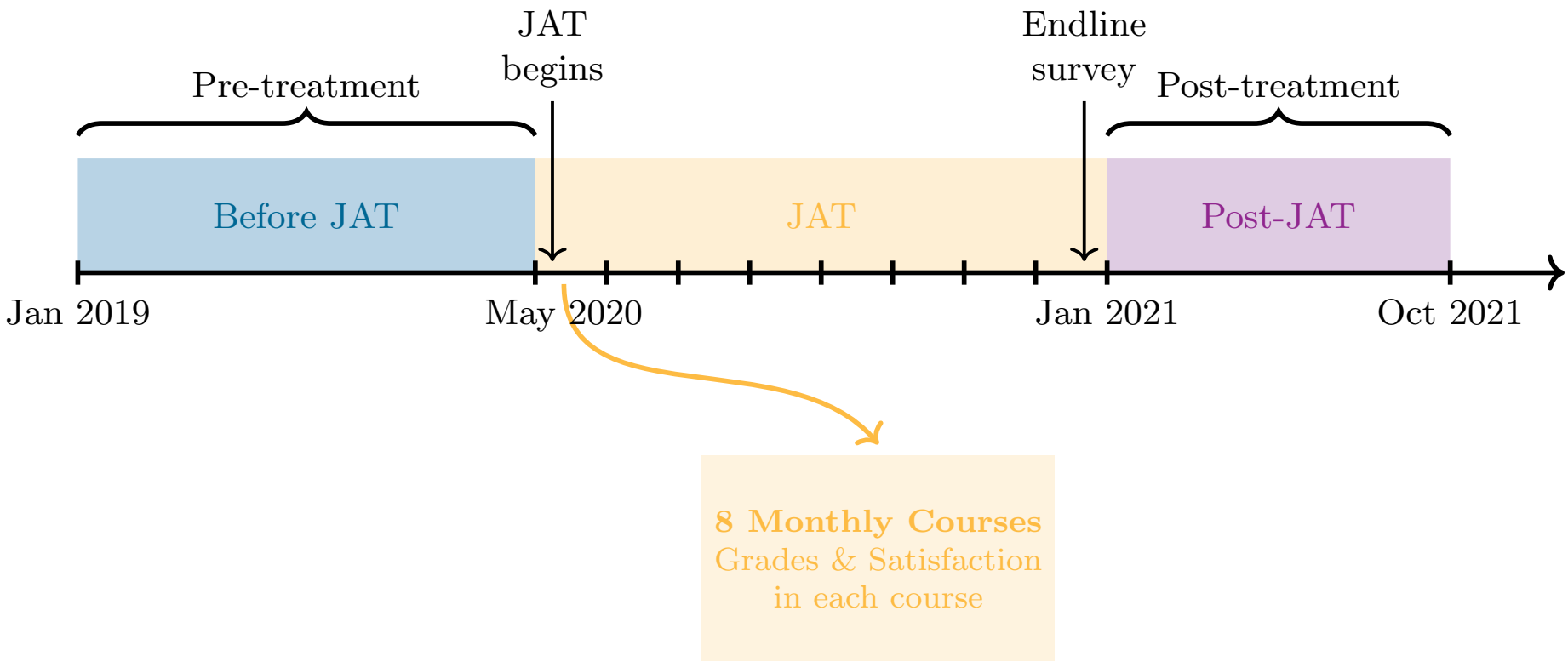

Figure: Program timeline

Program structure

Judicial Academy Training (JAT)								
May (Round 1)						June (Round 2)	...	Dec (Round 9)
Level	Total Students	Class #	Number of Students	Trainer Name	Monitoring Intervention			
2	157	1	32*	Trainer 1**	Yes (Mon A)***	...	...	...
		2	32	Trainer 2	No	...	...	...
		3	32	Trainer 3	Yes (Mon A)	...	...	...
		4	32	Trainer 4	No	...	...	...
3	457	5	25	Trainer 5	Yes (Mon B)	...	...	...
		6	25	Trainer 6	No	...	...	...
		...	...	...	...	...	...	...
		21	25	Trainer 21	Yes (Mon C)	...	...	...
		22	25	Trainer 22	No	...	...	...

*The numbers are based on a typical round. Some students did not take all classes, thus the total number of students is larger than the typical/average number of students in a class.

**Trainers change every round, each teaching 1 class per round up to 4 rounds.

*** There were 3 monitors in most rounds. 3 additional monitors assisted in some rounds.

Figure: Program structure

Course timeline

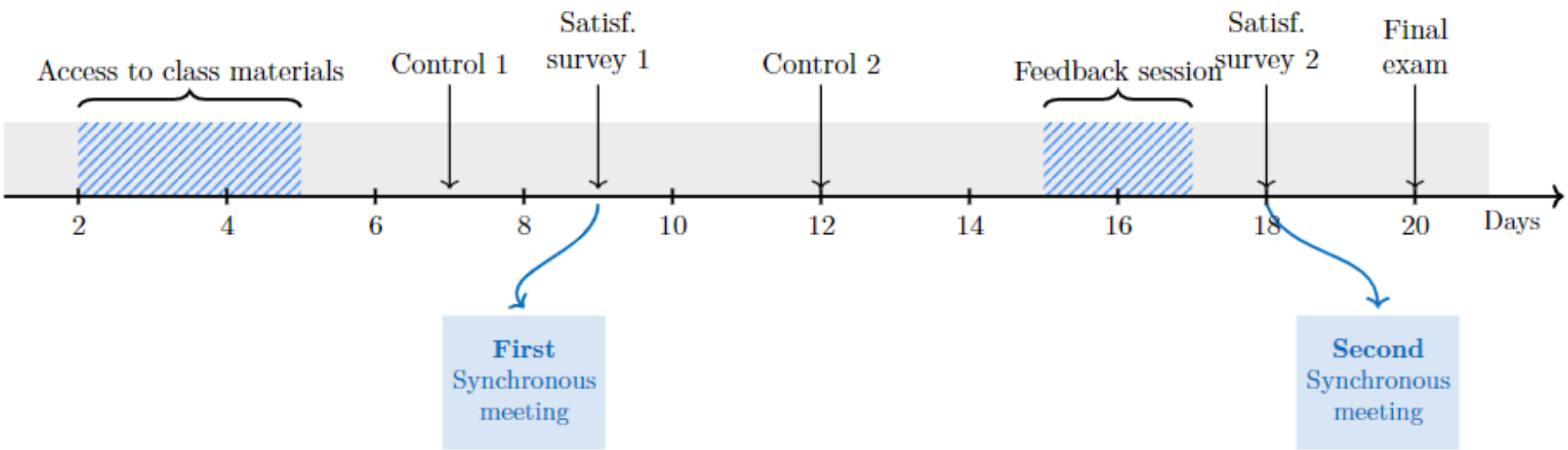

Figure: Course timeline

Study design

- Each monitor would make two monitoring visits per trainer's class (one hour for each visit divided into four fifteen-minute periods spread throughout the class)
- The monitors recorded their observations in a pre-specified observation form that evaluated the trainers on three criteria: teaching ability, mastery of the content, and key moments of the session
- After each monitored class, the monitor met the trainer for about 30 to 45 minutes to provide feedback and agree on areas of improvement
- Classes in the control group were not monitored and trainers did not receive any feedback

Table of Contents

Data

Data on academic and behavioral outcomes:

- Collected academic data from 9 rounds for each of the 604 students grouped in 22 classes. In total, 198 classes split into treatment and control.
- Data on grades includes students' grades in homework, tests, and final exams in each course
- Data on satisfaction includes AMAG's questionnaire on student satisfaction with the learning experience, trainer preparation, use of digital resources, and the training materials
- Survey data at endline: We ran a survey of behavioral exercises, which included a voluntary IAT to test gender-related implicit biases

Summary stats at the class level

Table 1. Summary statistics and balance table at the class level

Statistics	**Summary Statistics**						**Balance Tests**	
	Treatment			Control				
Class-level stats	Mean	Sd	N	Mean	Sd	N	β	p-value
Number of students	26.06	4.05	97	25.72	3.83	103	0.01	0.40
Share of female teachers	0.22	0.41	97	0.19	0.40	103	0.02	0.53
Share of judges	0.32	0.18	97	0.31	0.19	103	-0.23	0.45
Share of prosecutors	0.68	0.18	97	0.69	0.19	103	0.23	0.45
Share of female students	0.33	0.11	97	0.41	0.09	103	-1.99	0.23
Age	46.34	3.23	97	45.45	3.24	103	0.03	0.80
Years of tenure	5.30	0.78	97	5.40	0.61	103	-0.14	0.52
Years in the bar association	17.86	3.00	97	17.62	2.62	103	-0.02	0.87
Share in criminal court	0.17	0.11	97	0.17	0.12	103	0.17	0.54
Academy's specialist female	0.66	0.48	97					

Note: This table presents balance tests on the monitoring treatment. We present summary statistics displaying means and standard deviations for treatment classes ("Treatment") and control classes ("Control"). Balance tests present an OLS regression on treatment, with strata (participant level and location) and round fixed effects.

Data

Data on professional outcomes (case records):

- We scrape all publicly available case records for the 2018-2021 period. These include all non-criminal cases in Peru. We match each case record with the subset of judges enrolled in the JAT
- We create indicators to measure the efficiency and quality of the case resolution, such as:
  - clearance rate: ratio of cases resolved and cases filed
  - time to disposition: time from the filing to the resolution of the case
  - appeal rates: number of cases appealed out of the cases resolved
  - rates of appeals' reversal: rate of appeals that are reversed by the higher instance

Table of Contents

Empirical strategy: Grades

Specification to estimate the impact of the treatment on grades:

$$Grades_{icr} = \alpha + \beta Monitor_{cr} + \lambda_c + \eta_r + \gamma + \varepsilon_{icr}$$

Where $Grades_{icr}$ is a grade score for student i in course c during round r, $Monitor_{cr}$ is an indicator for receiving monitoring in course c during round r, λ_c and η_r are course and round fixed effects respectively and γ are strata controls. The coefficient of interest β indicates the average impact of the intervention on grades. Standard errors are adjusted for clustering at the class level.

Empirical strategy: Satisfaction

We proceed in a similar fashion to estimate the effects on satisfaction. Since there are 2 satisfaction surveys per round, we include m to account for the meeting:

$$Satisfaction_{imcr} = \alpha + \beta Monitor_{cr} + \lambda_c + \eta_r + \gamma + \varepsilon_{imcr}$$

Where $Satisfaction_{imcr}$ is a satisfaction measurement for student i in synchronous meeting m during course c and round r. The rest remains the same as for grades. Standard errors are adjusted for clustering at the class level too.

Results: Grades and Satisfaction

Treatment effects on grades and satisfaction

	Grades					Satisfaction	
	(1) Forum grade	(2) Reading grade	(3) Homework grade	(4) Exam grade	(5) Final grade	(6) With teacher	(7) With course
Monitoring	0.0702	0.0818**	0.0794	0.1609	0.1196**	0.0964*	0.0875*
	(0.0759)	(0.0347)	(0.0499)	(0.0956)	(0.0578)	(0.0553)	(0.0504)
Observations	4,968	4,988	5,017	5,000	5,021	10,023	9,967
R^2	0.13221	0.16559	0.12541	0.06765	0.09313	0.02617	0.03810
Dependent variable mean	0.04144	0.01453	0.05110	0.08771	0.07569	0.06086	0.06448
Round fixed effects	✓	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓	✓

Results: Grades and Satisfaction

Table: Monitoring effects on grades and satisfaction (no controls)

	Grades					Satisfaction	
	(1) Forum grade	(2) Reading grade	(3) Homework grade	(4) Exam grade	(5) Final grade	(6) With teacher	(7) With course
Monitoring	0.0629	0.0590	0.0884	0.1548	0.1238*	0.0964*	0.0971*
	(0.0762)	(0.0355)	(0.0542)	(0.0963)	(0.0603)	(0.0552)	(0.0531)
Observations	4,995	5,015	5,052	5,029	5,056	10,023	10,023
R^2	0.13727	0.15123	0.11348	0.06341	0.07999	0.02617	0.02996
Dependent variable mean	0.03987	0.01302	0.03900	0.08409	0.06193	0.06086	0.06466
Round fixed effects	✓	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓	✓

Note: Standard errors in parentheses, clustered at the class level. All columns include strata controls. Unit of observation is participant-round in columns 1-5 and participant-round-meeting in columns 6-7. Standard errors are clustered at the class level. Grades and satisfaction outcomes are standardized with respect to the control group mean. p ¡ 0.10, p ¡ 0.05, p ¡ 0.01

Results: Grades and satisfaction

- The intervention increases grades in the final exam by 0.12 SDs relative to the control group mean
- The intervention increases satisfaction with the trainer and with the course in 0.096 and 0.087 SDs
- In sum, the treatment improved both grades and satisfaction, especially for final grades

Empirical strategy: IAT scores

Specification to estimate effects on student's implicit gender stereotypes:

$$IAT_i = \alpha + \beta Monitor + \gamma + \varepsilon_i$$

Given selection into completing the IAT, we implement Lee Bounds to test if the results remain significant, both for the full sample and for each of the subsamples.

Results: IAT scores

Treatment effects on IAT scores

	Baseline			Baseline + Controls		
	(1) All	(2) Females	(3) Males	(4) All	(5) Females	(6) Males
Monitoring	0.3580**	0.1451	0.4183**	0.3575**	0.1362	0.4192**
	(0.1469)	(0.2268)	(0.1929)	(0.1498)	(0.2332)	(0.1957)
Lee Lower bound	-0.0065	-0.0571	-0.0057	-0.0065	-0.0571	-0.0057
Lee Upper bound	0.5551	0.2424	0.7446	0.5551	0.2424	0.7446
Observations	292	112	180	291	112	179
R^2	0.02836	0.07132	0.03628	0.03820	0.10496	0.06437
Dependent variable mean	0.15741	0.09413	0.19678	0.15607	0.09413	0.19482

Empirical strategy: Case outcomes

Specification to test the impact on judge performance indicators:

$$y_{im} = \alpha + \beta_1 Treatment_i + \beta_2 Y_{ci0} + \gamma + \varepsilon_{im}$$

Where y_{im} is the case outcome indicator for judge i in month m. $\beta Treatment_i$ is defined as the percentage of rounds treated over the duration of the PCA for judge i. Y_{ci0} is the pre-treatment mean of the dependent variable for twelve months prior to the start of the intervention, γ are strata controls. Standard errors are adjusted for clustering at the student level. We run this regression for outcomes measured after the intervention from January 2021 to October 2021.

Results: Case outcomes [Preliminary]

Treatment effects on case outcomes

	(1) Ruling favors plaintiff	(2) Appeal of ruling	(3) Reversal of ruling	(4) Clearance rate	(5) Time to disposition	(6) Timely Resolved
Panel A: Post						
Monitoring	0.0866	-0.1017	-0.0038	0.1683**	-0.2410	0.1799*
	(0.1189)	(0.1384)	(0.0591)	(0.0759)	(0.2485)	(0.1047)
Observations	169	169	169	203	219	219
R Squared	0.102	0.326	0.158	0.101	0.182	0.191
Dependent variable mean	0.8182	0.4915	0.0899	0.3220	-0.0496	0.4622
Panel B: DiD						
Monitoring	0.0451	0.1752	0.0661	0.3437***	-0.4401	0.0750
	(0.0727)	(0.1276)	(0.0774)	(0.0634)	(0.2714)	(0.1027)
Observations	548	548	548	781	788	788
R Squared	0.280	0.445	0.304	0.402	0.615	0.353
Dependent variable mean	0.8344	0.4593	0.1388	0.4792	-0.0277	0.5447

Table of Contents

Gender mechanism

We run regression specifications for grades, satisfaction and case outcomes splitting the sample by the participant's gender:

- The effects on academic scores and satisfaction seem to be driven by the female subsample

Gender mechanism

Treatment effects on grades and satisfaction by gender

	Grades					Satisfaction	
	(1) Forum grade	(2) Reading grade	(3) Homework grade	(4) Exam grade	(5) Final grade	(6) With teacher	(7) With course
Panel A: Males							
Monitoring	0.0556	0.0467	0.0536	0.1425	0.0829	0.0567	0.0332
	(0.0813)	(0.0470)	(0.0590)	(0.0950)	(0.0663)	(0.0592)	(0.0567)
Observations	3108	3123	3142	3129	3145	6248	6248
R Squared	0.137	0.162	0.119	0.057	0.088	0.035	0.042
Dependent variable mean	0.0371	-0.0076	0.0496	0.0836	0.0644	0.0518	0.0685
Panel B: Females							
Monitoring	0.0971	0.1437***	0.1012*	0.1769	0.1555**	0.1389	0.1794*
	(0.0734)	(0.0511)	(0.0518)	(0.1108)	(0.0672)	(0.0951)	(0.0969)
Observations	1860	1865	1875	1871	1876	3719	3719
R Squared	0.140	0.200	0.169	0.105	0.129	0.050	0.061
Dependent variable mean	0.0487	0.0516	0.0537	0.0945	0.0946	0.0747	0.0576

Gender mechanism

Table: Monitoring effects on grades and satisfaction

	Grades					Satisfaction	
	(1) Forum grade	(2) Reading grade	(3) Homework grade	(4) Exam grade	(5) Final grade	(6) With teacher	(7) With course
Monitoring	-0.2670	-0.3449	0.2506	0.4645*	0.0991	0.3618	0.3023
	(0.2179)	(0.2817)	(0.3296)	(0.2405)	(0.3343)	(0.3635)	(0.3198)
Female participant	-0.0182	-0.0279	-0.0275	-0.0126	-0.0305	-0.0279	-0.0848
	(0.0297)	(0.0414)	(0.0472)	(0.0380)	(0.0506)	(0.0977)	(0.1071)
Monitoring × Female participant	0.0642	0.1171	0.0739	0.0720	0.1300*	0.1266	0.1798
	(0.0413)	(0.0722)	(0.0663)	(0.0654)	(0.0706)	(0.1241)	(0.1290)
Monitoring × Participant age	0.0114*	0.0079	0.0008	0.0010	0.0076	-0.0011	0.0026
	(0.0063)	(0.0098)	(0.0091)	(0.0085)	(0.0112)	(0.0099)	(0.0090)
Monitoring × Participant judge	-0.2445**	0.1189	-0.1518	-0.0994	-0.0868	-0.2420**	-0.2342**
	(0.0954)	(0.1020)	(0.1086)	(0.1291)	(0.1412)	(0.1031)	(0.1027)
Monitoring × Years of experience	-0.0075	-0.0050	-0.0132	-0.0163	-0.0224	-0.0241	-0.0291*
	(0.0091)	(0.0124)	(0.0109)	(0.0116)	(0.0144)	(0.0164)	(0.0164)
Monitoring × Years of tenure	0.0023	7.04×10^{-5}	0.0126	-0.0119	0.0054	0.0400*	0.0278
	(0.0092)	(0.0115)	(0.0108)	(0.0089)	(0.0098)	(0.0216)	(0.0204)
Monitoring × Participant criminal court	0.1437*	0.1084	0.0485	0.0384	0.0904	0.0823	0.1691
	(0.0829)	(0.1294)	(0.1131)	(0.1156)	(0.1439)	(0.1805)	(0.1866)
Monitoring × Participant crim prosecutor	-0.1130*	0.1543*	-0.0863	0.0131	0.0198	0.0391	0.0483
	(0.0637)	(0.0895)	(0.0905)	(0.0740)	(0.0979)	(0.1215)	(0.1243)
Observations	4,968	4,988	5,017	5,000	5,021	9,967	9,967
R^2	0.13470	0.16945	0.12830	0.07187	0.09685	0.04236	0.04830
Dependent variable mean	0.04144	0.01453	0.05110	0.08771	0.07569	0.06036	0.06448
Round fixed effects	✓	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: Standard errors are clustered at the class level. Grades and satisfaction outcomes are standardized with respect to the control group mean.Regression includes criminal prosecutor/judge, age and experience controls, as well as interaction between controls and the treatment outcome. p ¡ 0.10, p ¡ 0.05, p ¡ 0.01.

Personalized feedback mechanism

To evaluate whether the effect is primarily driven by the monitoring or the feedback component of the intervention, we leverage the fact that the feedback only affects the second class, whereas the monitoring affects both classes

- We split the sample by first or second synchronous meeting, and include an additional variable with the average of the two meetings.
- We find that the effects seem to be driven by the second class, suggesting that the feedback component is driving the results of the intervention

Personalized feedback mechanism

Personalized feedback effects on satisfaction

	Satisfaction with teacher				Satisfaction with course			
	(1) First	(2) Second	(3) Avg.	(4) Diff.	(5) First	(6) Second	(7) Avg.	(8) Diff.
Monitoring	0.0482	0.1324**	0.0894*	0.0821***	0.0457	0.1295**	0.0866*	0.0814***
	(0.0527)	(0.0529)	(0.0518)	(0.0202)	(0.0502)	(0.0525)	(0.0505)	(0.0192)
Observations	4,986	4,981	4,988	4,979	4,986	4,981	4,988	4,979
R^2	0.03318	0.03584	0.03763	0.01182	0.04004	0.04111	0.04475	0.01431
Dependent variable mean	0.02951	0.09125	0.05933	0.05933	0.02765	0.10134	0.06341	0.07122
Round fixed effects	✓	✓	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓	✓	✓

Table of Contents

Next steps

- Working on the professional outcomes, both on the creation of indicators and the robustness of the results. The tables presented are at the judge-month level. Are they robust if we present them at the judge or the case level?
- How did trainers change their behavior? We could compare the notes from the first class to the notes from the second class we can see if the professor acted on the feedback. Perhaps do a qualitative assessment
- What are the satisfaction questions where we see the largest effects of the intervention, and are they related to potential mechanisms of the effect on gender?

- We could leverage a separate intervention (unrelated to this one) in which students were assigned to write self-reflection exercises to better understand the gender effects. Do students in the treatment group write about different topics (and, e.g., this might help explain the gender effects)? Could we find lower stereotypes in monitored classes over time?
- Are male or female students more likely to benefit from the first v. the second class?
- Evaluate any impacts on career promotions
- Other thoughts?

Table of Contents

Post regressions at the participant level

Overall

Dependent Variables:	Ruling favors plaintiff	Appeal of ruling	Reversal of ruling	Clearance rate	Days to resolution	Timely Resolved
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Monitoring	-0.0336	-0.1887	-0.0471	0.1791*	-111.7	0.2388
	(0.1447)	(0.1434)	(0.0790)	(0.1010)	(117.6)	(0.1525)
Fit statistics						
Observations	35	35	35	46	43	43
R^2	0.46684	0.64925	0.67672	0.54510	0.51779	0.57546
Dependent variable mean	0.83764	0.44579	0.08663	0.39467	222.43	0.59700

Clustered (Participant) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Weighted by case number

Dependent Variables:	Ruling favors plaintiff	Appeal of ruling	Reversal of ruling	Clearance rate	Days to resolution	Timely Resolved
Model:	(1)	(2)	(3)	(4)	(5)	(6)
Variables						
Monitoring	0.1211	-0.2726*	-0.1326	0.2024*	-158.5*	0.2100
	(0.0997)	(0.1505)	(0.0841)	(0.1015)	(82.92)	(0.1419)
Fit statistics						
Observations	35	35	35	46	43	43
R^2	0.76027	0.88001	0.87181	0.80538	0.72256	0.75202
Dependent variable mean	0.83764	0.44579	0.08663	0.39467	222.43	0.59700

Clustered (Participant) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

COURSE SCHEDULE

INDUCTION WEEK						
			Miercoles 14 de Oct. 20 20	Jueves 15 de otc. 2020	Viernes 16 de oct. 2020	Sbad 17 Oct 202
			START OF THE COURSE Revision of the guide didactics and syllabus of the course	Revision of the guide didactics, readings and case	Revision of the guide didactics, lectures and case	Review of material from study for the reading controls
Domingo 18 oct. 2020	**Lunes 19 oct. 2020**	**Martes 20 oct. 2020**	**Miércoles 21 oct. 2020**	**Jueves 22 oct. 2020**	**Viernes 23 oct. 2020**	**Sábado 24 Oct. 2020**
Review of material from study for the controls of reading	Video of Class N° 1 explanatory on the content of the course and component evaluative (45 minutes) Consultation chat N° 1 (45 minutes)	Review of material from study for the reading controls	Reading Control N° 1 (Start 00:00 hours)	Reading Control N° 1 (End 23:55 hours)	Video of Class N° 2 (Unit I and IL) (45 minutes) Consultation chat N° 2 (45 minutes)	1st. Synchronous session Unit I and II From 9:00 a.m. to 12:00 p.m. From 14:00 to 17:00

Figure: Program structure 1

Domingo 25 oct. 2020	Lunes 26 oct. 2020	Martes 27 oct. 2020	Miércoles 28 oct. 2020	Jueves 29 oct. 2020	Viernes 30 oct. 2020	Sábado31 Oct 2020
Review of material from study for the forum	 The discussion forum (Start 00:00 hours)	 The discussion forum (End 23:55 hours)	 Reading Control N° 2 (start 00:00 hours)	 Reading Control N° 2 (End 23:55 hours)	 Video of Class N°3 (Unit III) (45 minutes)  Consultation chat N° 3 (45 minutes)	2nd Synchronous session Unit III From 9:00 a.m. to 12:00 p.m.  From 14:00 to 17:00  Academic task
Domingo 1 nov. 2020	**Lunes 2 nov. 2020**	**Martes 3 nov. 2020**				
Review of material from study for the exam end	 Final exam (Start 00:00 hours)	 Final exam (End 23:55 hours) END OF THE COURSE				

Figure: Program structure 2

Table A4. Monitoring effects on grades and satisfaction

	Forum grade	Reading grade	Homework grade	Exam grade	Final grade	With teacher	With course
	Grades					Satisfaction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Monitoring	0.0583	0.0846**	0.0395	0.0950	0.0759	0.0820	0.0816
	(0.0853)	(0.0381)	(0.0449)	(0.1008)	(0.0572)	(0.0617)	(0.0583)
Monitoring × Participant level 2	0.0305	-0.1721***	0.3269***	0.4004***	0.3198***	0.0965	0.1037
	(0.1206)	(0.0486)	(0.1146)	(0.1211)	(0.0854)	(0.0913)	(0.1109)
Observations	4,995	5,015	5,052	5,029	5,056	10,023	10,023
R^2	0.13730	0.15227	0.11720	0.06893	0.08364	0.02648	0.03034
Dependent variable mean	0.03987	0.01302	0.03900	0.08409	0.06193	0.06086	0.06466
location_rand1 fixed effects	✓	✓	✓	✓	✓	✓	✓
Round fixed effects	✓	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: Standard errors are clustered at the class level. $* p < 0.10$, $** p < 0.05$, $*** p < 0.01$Grades and satisfaction outcomes are standardized with respect to the control group mean.Regression includes strata and round fixed effects

Figure: Heterogeneity by level

Table A6. Monitoring effects on grades and satisfaction by participant level

	Grades					Satisfaction	
	(1) Forum grade	(2) Reading grade	(3) Homework grade	(4) Exam grade	(5) Final grade	(6) With teacher	(7) With course
Panel A: Level 2							
Monitoring	0.1955***	-0.0498**	0.4996***	0.5008***	0.4801***	0.0706**	0.0277
	(0.0093)	(0.0116)	(0.0223)	(0.0226)	(0.0078)	(0.0229)	(0.0164)
Observations	1074	1084	1091	1089	1095	2172	2172
R Squared	0.278	0.061	0.186	0.216	0.146	0.076	0.080
Dependent variable mean	0.1431	0.0907	-0.0092	0.1309	0.0714	0.0419	0.0404
Panel B: Level 3							
Monitoring	0.0476	0.0844**	0.0292	0.0886	0.0686	0.0853	0.0847
	(0.0849)	(0.0387)	(0.0437)	(0.1000)	(0.0566)	(0.0612)	(0.0571)
Observations	3921	3931	3961	3940	3961	7851	7851
R Squared	0.088	0.167	0.086	0.032	0.062	0.014	0.019
Dependent variable mean	0.0116	-0.0084	0.0523	0.0711	0.0593	0.0661	0.0714

Note: Standard errors are clustered at the class level. Grades and satisfaction outcomes are standardized with respect to the control group mean. Panel A shows regression coefficients for the level 2 subsample. Panel B shows regression coefficients for the level 3 subsample. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure: Heterogeneity by level

Table A5. Monitoring effects on grades and satisfaction

	Forum grade	Reading grade	Homework grade	Exam grade	Final grade	With teacher	With course
	Grades					Satisfaction	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Monitoring	-0.2909	-0.1594	0.1243	0.2532	0.0357	0.5415	0.5413*
	(0.1844)	(0.2954)	(0.2831)	(0.2136)	(0.3064)	(0.3830)	(0.3065)
Monitoring × Participant level 2	0.0961	-0.1130	0.2362**	0.3348**	0.2108**	-0.0981	-0.1047
	(0.1122)	(0.0722)	(0.1020)	(0.1217)	(0.0935)	(0.1150)	(0.1184)
Monitoring × Participant age	0.0102	0.0047	-1.32×10^{-5}	0.0005	0.0053	-0.0045	-0.0019
	(0.0060)	(0.0095)	(0.0091)	(0.0081)	(0.0112)	(0.0104)	(0.0094)
Monitoring × Participant judge	-0.2383**	0.1388	-0.1496	-0.0995	-0.0750	-0.2205**	-0.2063*
	(0.0958)	(0.1045)	(0.1120)	(0.1308)	(0.1471)	(0.1043)	(0.1039)
Monitoring × Years of experience	-0.0032	-0.0045	-0.0052	-0.0056	-0.0133	-0.0229	-0.0273
	(0.0085)	(0.0121)	(0.0109)	(0.0100)	(0.0146)	(0.0176)	(0.0179)
Monitoring × Years of tenure	0.0036	0.0028	0.0140	-0.0106	0.0082	0.0429*	0.0313
	(0.0090)	(0.0106)	(0.0108)	(0.0085)	(0.0099)	(0.0217)	(0.0209)
Monitoring × Participant criminal court	0.1429	0.0569	0.0703	0.0771	0.0919	0.0310	0.1038
	(0.0873)	(0.1344)	(0.1260)	(0.1245)	(0.1651)	(0.1652)	(0.1706)
Monitoring × Participant crim prosecutor	-0.1106*	0.1537	-0.0821	0.0205	0.0254	0.0396	0.0458
	(0.0610)	(0.0904)	(0.0903)	(0.0690)	(0.0963)	(0.1197)	(0.1212)
Observations	4,968	4,988	5,017	5,000	5,021	9,967	9,967
R^2	0.13468	0.16884	0.12956	0.07445	0.09687	0.04152	0.04674
Dependent variable mean	0.04144	0.01453	0.05110	0.08771	0.07569	0.06036	0.06448
Round fixed effects	✓	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: Standard errors are clustered at the class level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$Grades and satisfaction outcomes are standardized with respect to the control group mean.Regression includes criminal prosecutor/judge, age and experience controls, as well as interaction between controls and the treatment outcome

Figure: Heterogeneity by level

Line of Training	Subject 2nd level	Subject 3rd level
Fundamental	**Course 1:** Fundamentals of Ethics in Judiciary	**Course 1:** legal argumentation
	Course 2: Constitutional Theory, Rights Fundamentals and Gender Approach in the administration of Justice	**Course 2:** Ethics in the Magistracy
	Course 3: Basic Fundamentals of Control conventionality and control constitutionality	**Course 3:** Control of Conventionality and Control of Constitutionality, Binding Precedents of the Constitutional Court and Standards of the IACHR
	Course 4: Interpretation and argument legal	**Course 4:** Emblematic cases of Law Constitutional Procedure
	Course 5: • Methodology for solving criminal cases based on the theory of criminal law and the legal consequences of the crime • Theory of judicial decision in civil matters.	**Course 5:** Constitutional Framework of Law Administrative
	Course 6: • Problems in the Civil Process • Debatable Institutions in the Criminal Process • Problems in the Administrative Litigation Process	**Course 6:** • Assessment of the evidence at the appeal venue in the Civil Process • Assessment of the evidence on appeal in the Criminal Process

specialized	• Problems in the Administrative Litigation Process	Assessment of the evidence on appeal in the Criminal Process
	Course 7: • Analysis of jurisprudence and plenary agreements of the Supreme Court of Justice in civil matters • Analysis of jurisprudence and plenary agreements of the Supreme Court of justice in criminal matters.	**Course 7:** • Civil challenge law • Criminal challenge law
	Course 8: • Emblematic cases of Family Law • forensic investigation • Emblematic cases of Administrative Law	**Course 8:** • Plenary agreements of the Supreme Court of Justice in Civil Matters • Plenary agreements of the Supreme Court of Justice in Criminal Matters
complementary	**Course 9:** • Management and leadership of the Tax Office • Management and leadership of the Judicial Office	**Course 9:** Public Management: Skills managerial
	Workshop: Oral litigation and direction of hearings Workshop: Methods and techniques of investigation and case theory Workshop: Interculturality in justice peruvian	Workshop: Problems in Oral Litigation and direction of hearings Workshop: Role of the prosecutor and the judge in the era digital

Rounds 1-5

Table: Monitoring effects on grades and satisfaction for first 5 rounds

	Grades					Satisfaction	
	(1) Forum grade	(2) Reading grade	(3) Homework grade	(4) Exam grade	(5) Final grade	(6) With teacher	(7) With course
Monitoring	0.0694	0.0404	0.0781	0.1075	0.0892	0.0679	0.0600
	(0.1005)	(0.0445)	(0.0873)	(0.1085)	(0.0818)	(0.0588)	(0.0570)
Observations	2,766	2,781	2,809	2,794	2,812	5,567	5,567
R^2	0.08168	0.17654	0.10586	0.02344	0.04838	0.02781	0.03338
Dependent variable mean	-0.03622	-0.09852	-0.01335	0.03666	-0.01458	0.06373	0.05448
Round fixed effects	✓	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓	✓

Note: Standard errors in parentheses, clustered at the class level. All columns include strata controls. Unit of observation is participant-round in columns 1-5 and participant-round-meeting in columns 6-7. Standard errors are clustered at the class level. Grades and satisfaction outcomes are standardized with respect to the control group mean. p ¡ 0.10, p ¡ 0.05, p ¡ 0.01

Rounds 6-9

Table: Monitoring effects on grades and satisfaction for last 4 rounds

	Grades					Satisfaction	
	(1) Forum grade	(2) Reading grade	(3) Homework grade	(4) Exam grade	(5) Final grade	(6) With teacher	(7) With course
Monitoring	0.0553	0.0711*	0.1357	0.2172*	0.1816**	0.1413*	0.1538**
	(0.0914)	(0.0413)	(0.0889)	(0.1208)	(0.0698)	(0.0769)	(0.0698)
Observations	2,229	2,234	2,243	2,235	2,244	4,456	4,456
R^2	0.21754	0.04982	0.13151	0.13520	0.13760	0.02816	0.02881
Dependent variable mean	0.13431	0.15186	0.10456	0.14338	0.15781	0.05728	0.07737
Round fixed effects	✓	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓	✓

Note: Standard errors in parentheses, clustered at the class level. All columns include strata controls. Unit of observation is participant-round in columns 1-5 and participant-round-meeting in columns 6-7. Standard errors are clustered at the class level. Grades and satisfaction outcomes are standardized with respect to the control group mean. p ¡ 0.10, p ¡ 0.05, p ¡ 0.01

Summary stats at the student level

Table 2. Summary statistics and balance table

Statistics	Summary Statistics						Balance Tests	
	Treatment			Control				
	Mean	Sd	N	Mean	Sd	N	β	p-value
Full sample								
Participant female	0.351	0.478	302	0.417	0.494	302	-0.078	0.031
Participant age	46.024	7.071	295	45.469	6.693	290	0.002	0.416
Participant judge	0.323	0.468	294	0.297	0.458	296	0.016	0.675
Years of tenure	5.216	1.780	287	5.493	2.765	284	-0.007	0.262
Years in the bar association	17.544	5.650	287	17.768	5.318	284	-0.001	0.776
Criminal court	0.522	0.502	92	0.552	0.500	87	-0.047	0.487
Criminal Prosecutor's Office	0.533	0.500	195	0.473	0.500	201	-0.001	0.989
Only judges								
Participant female	0.358	0.482	95	0.420	0.496	88	-0.063	0.340
Participant age	47.383	6.239	94	47.477	5.556	88	-0.003	0.621
Years of tenure	5.576	1.923	92	6.046	3.560	87	-0.003	0.690
Years in the bar association	18.978	5.116	92	19.782	5.054	87	-0.004	0.511
Criminal court	0.522	0.502	92	0.552	0.500	87	-0.047	0.487
Only judges with cases								
Participant female	0.531	0.507	32	0.645	0.486	31	-0.082	0.380
Participant age	46.161	6.558	31	46.774	5.420	31	-0.012	0.173
Years of tenure	5.276	0.841	29	5.448	1.088	29	0.024	0.546
Years in the bar association	17.897	5.287	29	19.345	4.418	29	-0.013	0.202
Pre-treatment case outcomes								
Days to resolution	145.549	141.397	244	144.545	126.293	295	0.000	0.495
Case timely resolved	0.595	0.392	244	0.574	0.376	295	0.028	0.460
Clearance rate	0.421	0.332	278	0.664	0.313	277	-0.152	0.009

Note: This table reflects balance tests on the monitoring treatment. We present summary statistics displaying means and standard deviations for the sample in the treatment classes ("Treatment") and for the sample in the control classes ("Control"). For the purpose of the summary statistics, the treatment is classified as attending at least 50% of treated classes, and the opposite for the control group