Bureaucratic Training and State Capacity: Experimental Evidence from Peru's Judiciary *

Daniel Chen[†] Manuel Ramos-Maqueda[‡] Bernardo Silveira [§]

April 30, 2026

Abstract

We employ a randomized controlled trial to evaluate an intervention aimed at improving the quality of in-service training of public servants in the Peruvian judiciary. The intervention involves pedagogical specialists who remotely monitor, assess, and provide feedback to instructors of courses taught to active judges and prosecutors in the country's judicial academy. We find that the intervention significantly improves educational outcomes, including grades and students' satisfaction with the course. Furthermore, the intervention improves the quality of judicial proceedings of treated judges, increasing the ratio of verdicts to total decisions and the rate of judges' attendance to litigant requests. This study demonstrates that a simple, scalable virtual intervention can effectively improve the training of bureaucrats and, consequently, enhance the quality of public service delivery.

*We would like to thank the Judicial Academy of Peru, and in particular, the current Academic Director, Nathalie Ingaruca Ruiz, and the former Academic Director, Hipolito Rodriguez Casavilca, for their invaluable partnership and support in this research project. We also acknowledge the exceptional efforts of the Judicial Academy team that implemented the intervention, with special thanks to Janina Toro, Sylvana Vasquez, Juan Carlos Del Carpio, Carlo Pajuelo, and Mercedes Dueñas. Our thanks go further to World Bank Project Team Leads, Carolina Vaira and Nicholas Menzies, as well as to Brandon Mora for his research support. We are grateful to Clare Leaver and Kate Orkin for their comments and guidance, as well as to Natalie Bau, Dennis Egger, Julien Labonne, Anandi Mani, Christian Meyer, Niclas Moneke, Christopher Woodruff, and participants of seminars at UC Berkeley, Oxford, and UC Riverside. Nithya Nagarathinam and Viknesh Nagarathinam provided excellent research support. This work benefited from funding from the World Bank Knowledge for Change (KCP) and DIME/i2i programs. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

[†]Toulouse School of Economics; email: daniel.chen@iast.fr

[‡]University of Oxford; The World Bank; email: mramosmaqueda@worldbank.org

[§]University of California, Los Angeles; email: silveira@econ.ucla.edu

1 Introduction

Professional training for public officials is ubiquitous. Every year, countries around the world provide in-service training to police officers, judges, procurement and tax officials, doctors, nurses, teachers, and social workers who deliver critical state functions. This investment is substantial: in 2017 alone, the U.S. spent approximately $10 billion training civil servants, representing roughly 4% of annual federal and state personnel budgets (Credential Engine, 2021). Despite this scale, research on how to improve the efficacy of such training remains sparse. The few existing studies evaluate the addition of new *content* to civil servants' curricula (Azulai et al., 2020; Banerjee et al., 2021; Mehmood et al., 2024).

Evidence from school-based education suggests that the delivery margin can be consequential. A large literature demonstrates that interventions targeting *how* teachers teach—through classroom observation, coaching, and structured feedback—can substantially improve student learning (Glewwe and Muralidharan, 2016; Kremer et al., 2013; Evans and Popova, 2016; Allen et al., 2011). Whether similar approaches work in the context of professional civil servant training is an open question. Civil servants differ from schoolchildren in ways that may limit the transportability of these findings: they bring decades of domain expertise to the classroom, they are taught by practitioners rather than career educators, and the institutional environments in which they train often lack basic pedagogical infrastructure. Whether structured pedagogical support can overcome these constraints—and translate into measurable changes in official conduct—remains unknown.

In this paper, we provide what is, to our knowledge, the first experimental evidence that improving instructor pedagogy can enhance the training outcomes—and, ultimately, the job performance—of civil servants.[1] We partner with the Judicial Academy of Peru (AMAG), the country's premier institution for the instruction of judges and prosecutors, to experimentally evaluate an intervention that targets instructor delivery quality directly. The intervention consisted of remotely monitoring, assessing, and providing feedback to instructors of an eight-month training program for active judges and prosecutors seeking promotion. We find that the intervention significantly improved the educational outcomes of class participants, leading to higher grades and satisfaction with the course. More importantly, it raised the quality of judicial service delivery by treated judges in the year after the program,

[1] This is also, to our knowledge, the first randomized controlled trial to evaluate a judicial training intervention anywhere in the world.

increasing both the ratio of cases reaching a verdict and the ratio at which judges attend litigants' requests.

The judicial setting in Peru is particularly well suited for this research, for two reasons. First, the judiciary is a cornerstone of state capacity, given its crucial role in contract enforcement and protection of citizen rights (Amirapu, 2021; Ramos-Maqueda and Chen, 2021; Kondylis and Stein, 2023; Chakraborty, 2016; Chemin, 2012; Laeven and Woodruff, 2007; Acemoglu and Johnson, 2005). Yet, as in many other developing countries, the judiciary in Peru is often perceived as inaccessible, unreliable, and inefficient. In a 2018 survey of Peruvian citizens, 68% of the respondents believed that most or all judges are corrupt, while 63% felt the same about prosecutors. Only 13% of the respondents reported having some trust in judges and magistrates, with the corresponding number for prosecutors being 16% (World Justice Project, 2022). Furthermore, the judiciary often faces weak professional standards, which have been associated with weakness in the training of court actors (Romero Osorio, 2020).

Second, AMAG exemplifies the instructional challenges that motivate our intervention. The Academy has been chronically under-funded, instructors are hired on a course-by-course basis with no pedagogical training requirements, courses receive minimal classroom observation and no structured feedback on teaching quality, and instruction has historically defaulted to expository lectures with limited active learning—despite the practical nature of judicial work. For experienced professionals managing heavy workloads, this format offers limited actionable learning and, consequently, limited potential to change behavior on the job. (Section 2.2 provides further institutional detail.) These challenges were compounded in 2020 when the program transitioned to online delivery in response to the COVID-19 pandemic, exposing gaps in instructors' ability to engage students in virtual formats. While our evidence comes from AMAG, these features—practitioner instructors hired on short-term contracts, minimal pedagogical infrastructure, and a reliance on expository lecture—are also documented across judicial academies in Latin America (González and Cooper, 2017), and extend to civil service training more broadly, where subject-matter expertise and teaching quality are rarely developed in tandem (Schaffner et al., 2024).

Our intervention included 604 active judges and prosecutors enrolled in an online training program—AMAG's Promotion Program (APP)—from May to December 2020. The training consisted of nine consecutive courses of four weeks each. These courses covered various disciplines, from fundamental topics (e.g., ethics, judicial and legal interpretation), to specialized law subjects (e.g., civil, criminal, and administra-

tive law), and complementary topics (e.g., managing judicial offices and case-based theory). Students were divided into 22 classes based on their professional position and geographical location. Within each course, half of the classes were randomly assigned to receive the treatment, while the rest were assigned to the control group. The intervention followed closely the Pre-Analysis Plan registered in the AEA RCT Registry.[2]

At the beginning of each course, AMAG informed instructors in treated classes about the intervention, consisting of the monitoring, assessment and provision of feedback to instructors. Specifically, a pedagogical specialist would attend parts of the first session of online lectures and assess the instructor's class delivery. Following the first session, the specialist would meet privately with the instructor to review the delivery of the class and offer feedback, covering topics such as student participation, case-based teaching, class organization and management, and the effective use of technology. The specialist would then join the second session of online lectures, scheduled for two weeks after the first one, and assess the instructor again. Instructors in control group classes did not receive any monitoring, assessment, or mid-course feedback. Furthermore, students were not informed of the intervention or of the existence of treated and control classes. All classes in our study—regardless of treatment assignment—followed the same structure and format as any other class in the Judicial Academy.

The intervention improved the educational outcomes of students in treated classes relative to the control group. We observe an increase of approximately 0.22 standard deviations in the final grades of treated students, with statistically significant gains also evident in reading, homework, and exam grades.[3] Tests and exams were identical across treatment and control classes within each course, and instructors followed standardized grading criteria, facilitating direct comparability. Grades in the program are meaningful to students, as they constitute an explicit component in judiciary promotion decisions; Section 2.1 provides more details. Students in treated classes also reported significantly higher satisfaction with both the instructor and the course.

[2]The pre-analysis plan (PAP) can be found online at https://doi.org/10.1257/rct.7113-1.0. This PAP encompasses two distinct interventions, although they are documented under the same registration. The current paper exclusively examines the first intervention, referred to in the PAP as a "teacher monitoring" intervention. IRB approval was obtained from Solutions IRB (ID: 0193).

[3]Because some instructors teach in more than one round of the program, our main classroom analysis restricts the sample to each instructor's first teaching episode in the APP program in 2020, isolating the treatment-control contrast from any cross-round carry-over in instructor behavior. See Section 3 for details.

We then assess whether the treatment led to improved judicial service delivery outside the classroom. We collect data on judges' professional performance in 2021, the year following the intervention, and match this data with judges who participated in the 2020 training program.[4] Due to the lack of publicly available data for prosecutors, this part of our analysis focuses solely on judges. Based on the available data, we build indicators of the productivity and the quality of judicial processes.[5] First, we find that treated judges are relatively more likely to reach verdicts in their cases rather than to end them for procedural reasons at earlier stages—which we interpret as evidence that our intervention fostered more thorough consideration of the merits of the case. This improvement in the quality of judicial processes comes at no expense of productivity, as the treatment has no impact on the judges' clearance rate (the ratio of cases resolved over cases filed).

Additionally, we evaluate the impact of the intervention on the judges' engagement in a recently implemented program called "The Judge Listens to You," through which litigants can request appointments with their judges. This program aims to bring judges closer to citizens, improving their sense of proximity and access to the justice system. We find that treated judges substantially increase the likelihood of attending meetings requested by the litigants, reducing the no-show rates. After the intervention, treated judges increased their attendance rate by 26.9 percentage points relative to a control-group mean of 73 percent.. These results point towards a more attentive and better service for citizens as a result of the treatment.

Given the importance of these results, we examine whether the improvements are more consistent with the monitoring or feedback component of the intervention. Monitoring is the process whereby the specialist observes and evaluates the instructor's performance during both the first and the second class sessions. Feedback is provided to the instructor after the first session and includes a written commitment to improving their delivery in the second class. Thus, while the monitoring component affects instruction in both sessions, feedback specifically targets improvements in the latter session. Although there is no clean causal design to separately identify the impact of the monitoring and feedback components, we would expect to see similar

[4]The data comes from the publicly available website "Conoce a tu juez", which includes productivity metrics of judges: https://sap.pj.gob.pe/casillero-digital/#/conoce-tu-juez.

[5]As explained in greater detail in section 3, we rely on publicly available data to measure judicial performance, which allows us to build a limited set of performance indicators. While the indicators we can evaluate with this data do not exactly match those in the PAP, they still allow us to measure the impact of the intervention on both productivity and quality metrics, which is the original intention described in the PAP.

effects on class outcomes in both sessions if monitoring were the key mechanism for improved teaching. Conversely, if feedback were the main driver, we would expect improvements in the second session compared to the first.

The evidence is more consistent with the latter pattern. Treatment had no significant effect on student satisfaction in the first session, but produced a substantial increase in the second session. In addition, AMAG specialists' evaluations of treated instructors improve from the first to the second session across all three measured dimensions: teaching skills, content mastery, and class structure. We interpret these patterns as suggestive evidence that feedback was an important channel through which the intervention improved instructional quality. This comparison does not separately identify the two components causally; for instance, monitoring itself may have dynamic effects, and repeated observation could have induced instructors to adjust their behavior over time. We therefore view the evidence as consistent with a meaningful role for feedback, while acknowledging that other mechanisms may be at play.

Our study contributes to the literature in several ways. First, and most broadly, we contribute to the burgeoning literature on the in-service training of civil servants. As evidenced by its omission in previous literature reviews, this type of training has, until recently, received limited attention from empirical research on the personnel economics of the state (Finan et al., 2017; Hansen and Tummers, 2020)—a gap that is especially striking given the evidence of large dispersion in bureaucratic effectiveness within government organizations, which points to the need for interventions targeting lower-performing public employees (Best et al., 2023). Three recent papers have contributed to filling this gap, showcasing how newly designed training modules can improve civil servants' performance: Banerjee et al. (2021) show that a three-day soft-skills training enhanced the performance of police officers in Rajasthan, India; Azulai et al. (2020) evaluate a new productivity training module delivered through Ghana's Civil Service; and Mehmood et al. (2024) evaluate an empathy training delivered to junior ministers in Pakistan. In contrast, our contribution is to examine the *delivery* of existing instruction, drawing on a literature on school-age education that identifies pedagogical interventions as among the most effective levers for improving student outcomes (Kremer et al., 2013; Araujo et al., 2016; Glewwe and Muralidharan, 2016; Evans and Popova, 2016; Piper et al., 2018; Angrist et al., 2024).

Second, our study is also, to the best of our knowledge, the first randomized controlled trial on judicial training, as well as the first empirical paper to study judicial

training in a developing country context. This is particularly relevant given the slow adoption of empirical methods in the legal field. The few existing studies on this subject have focused on the United States, evaluating the integration of new topics into judges' curriculum (Ash et al., 2022; Baye and Wright, 2011). We instead focus on improving a training program provided by an under-resourced educational institution in Peru, a developing country marked by low confidence in the judiciary by the general public. Over the years, international organizations have invested substantial amounts in capacity-building programs aimed at improving the training of judicial actors and strengthening the rule of law across developing countries: USAID has historically allocated 25% of its justice reform assistance to judicial training (Hammergren, 1998), while over 50% of The World Bank's lending operations on justice reform have supported training programs (The World Bank, 2012). Despite such efforts and investments, there is very limited evidence on effective training interventions to improve judicial performance. Ours is one of the few empirical studies to date showcasing the potential of simple, cost-effective interventions to improve the performance of justice systems throughout the world (Sadka et al., 2024; Chemin et al., 2022).

Finally, our intervention is embedded in the concept of "Community of Practice" (CoP), which underscores the importance of shared learning, social interaction, and collective knowledge development in professional settings. Since the seminal article by Wenger (1998) introduced the concept, CoPs have been influential in the organizational and pedagogical literature and practice. A CoP is defined as a group of people who "share an interest for something they do and learn how to do it better as they interact regularly" (Wenger, 2011). Despite being widely used in bureaucracies (Cuddy, 2002), CoPs have seen limited randomized controlled trial evidence. We provide, to the best of our knowledge, the first experimental evidence evaluating the impact of CoP-inspired pedagogical interventions in the civil service context, and in so doing integrate CoP principles into the economics of governance and bureaucratic performance.

The remainder of the paper proceeds as follows. Section 2 discusses the background context and the study design. Section 3 presents the data sources and outcomes of interest. We then present our empirical results in Section 4. Section 5 concludes.

2 Context and Study Design

2.1 Context: The Judicial Academy Training

AMAG is a government agency in Peru dedicated to providing training to active and aspiring judges, prosecutors, and clerks. Each year, the Academy offers education to over 10,000 legal professionals. In this study, we partner with AMAG's Promotion Program (APP), which trains active judges and prosecutors seeking career advancement. The APP is a critical component of the promotion process for Peruvian judges and prosecutors. Completion of the APP is required for candidates to qualify for advancement, and the grade obtained in the program influences the promotion outcomes. Specifically, promotion decisions are partly based on a numerical score that depends on the applicant's APP grade.[6]

The APP offers three tiers of training tailored to the varying experience levels of legal professionals. These tiers correspond to promotion courses for levels 2, 3, and 4, aligned with the hierarchical positions of judges and prosecutors. Level 2 applicants include judges in lower courts and adjunct or assistant provincial prosecutors who aim to become specialized judges or provincial prosecutors. Level 3 applicants are specialized judges or provincial prosecutors aspiring to positions in high-level regional courts or public prosecution offices. Finally, level 4 applicants are high-level regional judges seeking promotion to the Supreme Court. As there was only one class of level 4 applicants in the period covered by our intervention, we exclude it from this study.

The 22nd APP was conducted from May to December 2020, offering nine courses (rounds) of four weeks for each level. The nine courses each cover a separate topic, categorized into three groups: fundamental topics (courses 1-4), specialized topics (courses 5-8), and a complementary course. Fundamental courses addressed ethics in the judiciary, legal interpretation and argumentation, constitutional law and appeals. Specialized topics split students into areas of law based on their specialty—generally civil or criminal law, with specific subspecialties such as family and administrative law. These courses cover methodology for resolving cases, problems faced in the judicial proceedings, jurisprudence in these fields, or emblematic cases, among others. The complementary course dealt with judicial office management and included a two-week elective workshop. In addition, each of these courses took place in the midst of

[6]As per the Regulation for the Promotion of Judges and Prosecutors in Peru, the promotion process will consider the applicant's grade in the APP, "if the applicant has obtained a grade equal to or greater than 13 in the program that corresponds to the level to which she aspires" (Reglamento de Concursos para el Ascenso de Jueces y Fiscales, Junta Nacional de Justicia.)

a transition towards case-based teaching in the Academy, with increased emphasis on practical applications of the lessons into the judge and prosecutors' day-to-day jobs. More detailed information on the topics of the program can be found in Appendix B.

Each of these courses lasts approximately four weeks. In the initial days of the course, participants are granted online access to class materials. Following the review of these materials, students have to complete online activities and assignments. They also attend day-long sessions on the second and fourth Saturday of each course, spanning from 9 a.m. to 5 p.m. At the end of each of these day-long sessions, students are required to fill out the satisfaction survey provided by the Judicial Academy. Throughout the course, students complete two tests and a final examination. More detailed information on a typical course schedule can be found in Appendix C.

The instructors of each course are legal practitioners and university professors with a specialization in a legal discipline and a minimum of four years of postgraduate experience. They are hired by the Judicial Academy to teach a specific course related to their specialization, so the assigned instructor typically changes with each course, and few teach more than one course per program. In accordance with AMAG's formal requirements, instructors must hold a master's or doctoral degree and qualify under one of several categories: judges or prosecutors with at least four years of service, university professors with at least four years (or eight semesters) of teaching experience, research professors with at least four publications in the past ten years, or foreign professors with at least four years of experience in their field.

2.2 Challenges in the Quality of Instruction

As discussed in the introduction, a key premise of our intervention is that instructor delivery quality is a crucial constraint on the effectiveness of civil servant training. In this subsection, we document the specific institutional features of AMAG that support this premise. While the Academy plays a central role in judicial career advancement, several structural factors have historically limited the quality of classroom instruction. The institution has been chronically under-funded, receiving only approximately 0.2% of the justice sector budget according to OECD reviews. This severe resource constraint has limited AMAG's ability to hire full-time faculty, expand programs, or provide strategic direction for training needs (World Bank, 2004). Consequently, both the quality of instructional programs and the Academy's capacity to attract and retain high-quality instructors have suffered.

These institutional weaknesses manifest directly in the quality of classroom in-

struction. First, courses have historically relied on expository lectures with limited case-based discussion or other active-learning components, despite the practical nature of judicial work. Second, instructors are typically legal experts recruited on a course-by-course basis who receive little systematic classroom observation or formative feedback on their teaching. Third, because instructors rotate frequently and are not trained for continuous instruction, the quality of teaching varies substantially across courses and rounds—including differences in how instructors organize their sessions and engage students. External reviews by the Justice Reform Commission and organizations such as Justicia Viva have repeatedly highlighted these shortcomings, noting the lack of innovative teaching techniques, insufficient evaluation of outcomes, and minimal accountability for underperforming instructors. Evaluation of instructors relied primarily on student satisfaction surveys collected at end of session, providing little timely, formative input on which instructors could act mid-course to enhance their teaching.

These challenges were compounded in 2020 when AMAG transitioned to online training in response to the COVID-19 pandemic. Many instructors were new to virtual formats, which presented additional obstacles to effective class delivery. The shift to online instruction increased the returns to clear session design and effective use of digital tools—precisely the areas where instructors had limited structured support. These features likely widened the gap between what was taught in APP courses and the practical competencies that matter on the bench, such as structured legal argumentation, issue spotting, case analysis, and responsiveness to litigants. This gap was particularly consequential given the high stakes of APP grades for career progression. The combination of resource constraints, inconsistent instructional quality, limited instructor support, and the abrupt transition to online learning created a strong rationale for implementing a systematic approach to instructor observation and feedback.

2.3 Intervention: Instructor Monitoring and Feedback Program

To address these constraints, AMAG rolled out an innovative intervention across the entire APP program from May to December 2020. The intervention was randomized at the class level. Specifically, for each four-week long course, students were divided into 22 classes—with four classes for level 2 students and 18 classes for students in level 3. In the first course of the program, the 22 classes were randomly assigned into

treatment and control groups, with stratification based on class level and location.[7] Thus, 11 classes—each with a different instructor—were treated, while the remaining 11 were in the control group. Students remained in the same treatment group during the majority of the courses, with some exceptions when they selected specialized courses in rounds 6-8. Out of the 604 students in total, 456 (representing 75.5% of the total) were fully treated or never treated, while the remaining students participated in some treated classes but not in all of them (see Figure A1 in the Appendix).[8] [9]

Throughout the program, 97 classes received the treatment and 103 classes were in the control group.[10] Treated classes had instructors who were observed, assessed, and received feedback by a specialist from the Academy, while control classes did not receive any observation or feedback.

Each treated class had a unique instructor who was assigned a specialist. The specialist was a pedagogical expert already employed by the Judicial Academy, responsible for implementing the intervention in up to four classes per course. There were a total of six different specialists in charge of implementing the intervention throughout the entire program. These specialists also had prior teaching experience and involvement in guiding and monitoring instructional performance. They had been trained in constructive communication skills, enabling the monitored instructors to receive feedback positively and to facilitate improvements in their teaching processes.

The intervention consisted of two main components: instructor observation and assessment, and subsequent feedback to enhance performance. The pedagogical team at AMAG designed class evaluation guidelines, which included an "Observation Form" for standardized assessment of instructors, and an "Agreements and Commitments Form" for standardized feedback sessions. The latter form ensured that the discussion between the specialist and the instructor resulted in a set of agreements and commitments that the instructor would have to consider in their next class. The intervention was designed by methodological specialists at AMAG, based on evidence from the Education literature and on a review of similar experiences in other settings.

[7] While the instruction was online, students were grouped into different classes based on their location. This is why we also use class location as strata.

[8] Among participants with any treatment exposure, the mean share of treated courses is 0.795. The implied treatment-on-the-treated effect is therefore only modestly larger than the ITT benchmark ($1/0.795 \approx 1.26$), suggesting that the results do not require a large compliance adjustment to be meaningful.

[9] Appendix Table A6 shows that the share of treated participants is approximately 50% across all courses, confirming that randomization achieved balance by courses as well as overall.

[10] For ease of understanding, Figure 1 summarizes the structure of the program and the treatment assignment across classes and rounds.

The timeline of the intervention for each course is as follows: the specialist first assessed the delivery of the class in the first class session, following the criteria in the "Observation Form." After the first session, the specialist provided feedback to the instructor following the "Agreements and Commitments Form." Finally, two weeks after the first class session, the specialist assessed the delivery of the second session, following the same "Observation form" as in the first session. This timeline of events is summarized in Figure 2, which highlights in red the key components of the intervention.

During the first session, the specialist visited each class four times throughout the day, with each visit lasting an average of 15 minutes. Thus, the specialist spent a total of one hour in each class every day. During the visits, the specialist took notes, recorded in the observation form, which assessed the instructor in three main areas: (i) *teaching skills* (e.g., enthusiasm, student relationships, voice modulation, and language clarity); (ii) *content mastery* (e.g., the clarity of the presentation and the use of real life examples); and (iii) *class structure* (e.g., the use of motivational strategies and the evaluation of students). Appendix D provides details of the assessment indicators in the observation form.

In the week following the first session of classes, the specialist and instructor had a 30-minute feedback meeting to discuss and offer feedback on the teaching experience during the first session. During this meeting, the specialist first asked the instructor to perform a self-assessment, including their assessment of their own strengths and areas for improvement in the teaching skills, mastery of the content, and class structure. The specialist then shared their own feedback in terms of both the strengths and areas for improvement in those same topics. Based on this discussion, the instructor and specialist reached a set of agreements and commitments on how to improve the delivery of the class in the upcoming session of classes. All of this was documented in the "Agreements and Commitments Form," which was signed by both the specialist and the instructor. A sample of this form can be found in Appendix E.

3 Data

3.1 Academic performance and student satisfaction

There were 604 judges and prosecutors who enrolled in the APP. Most students took level 3 (72.77%), followed by level 2 (22.93%), while only 4.30% took level 4. As previously explained, we exclude level 4 from this study, since it had only one

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 1. Program structure

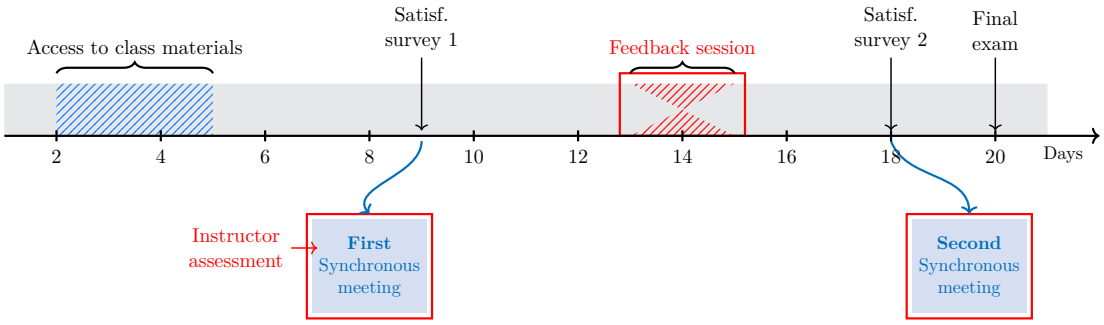

Figure 2. PCA - Calendar course

class per course. The proportions of prosecutors in levels 2 and 3 are 70.83% and 64.99%, respectively, with the rest of the participants being judges. Table 1 shows the summary statistics and balance tests on participant and class characteristics.

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

	Summary Statistics						Balance Tests	
	Treatment			Control				
Statistics	Mean	Sd	N	Mean	Sd	N	β	p-value
Number of students	26.062	4.054	97	25.718	3.828	103	-0.002	0.899
Share of female teachers	0.216	0.414	97	0.194	0.398	103	-0.044	0.615
Share of judges	0.317	0.180	97	0.311	0.193	103	0.015	0.770
Share of prosecutors	0.683	0.180	97	0.689	0.193	103	-0.015	0.770
Share of female students	0.334	0.108	97	0.408	0.090	103	-0.078	0.032
Age	46.341	3.233	97	45.447	3.240	103	0.003	0.306
Years of tenure	5.304	0.783	97	5.402	0.612	103	-0.007	0.422
Years in the bar association	17.858	3.003	97	17.621	2.623	103	-0.001	0.789
Share in criminal court	0.168	0.111	97	0.169	0.121	103	-0.006	0.874
Academy's specialist female	0.663	0.475	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 of treatment on each characteristic, controlling for strata (participant level and location) and round fixed effects. The regression for number of students is estimated at the class level (N=22) with HC1 robust standard errors. All other variables are estimated at the individual level with standard errors clustered by class. The monitor's gender (Academy's specialist female) appears in the summary statistics columns only and is excluded from balance regressions, as only treated classes have a monitor assigned.

Administrative data on grades and student satisfaction was collected over nine rounds for all judges and prosecutors enrolled in the program. The grade data includes detailed information on homework, tests, and final exam grades for each round. These tests and exams were the same for all students in the same course, regardless of the class treatment status. The satisfaction data consists of a series of Likert-scale questions about the student's satisfaction with their learning experience, the instructor's preparation, and the use of digital resources, among others. Students answered satisfaction-related questions twice within a round, right after each day-long class, and their participation in this survey is required by the Academy. These satisfaction questions are collected by the Academy before the final exam and before students receive their final grades in the course.

Some instructors teach more than one round during the program; see Appendix Table A7 for details. Because exposure to monitoring and feedback may induce persistent changes in instructor behavior, our main classroom analysis uses class-round combinations corresponding to each instructor's first time teaching in the program.

We report full-sample estimates in the Appendix as a robustness check. We also report summary statistics and balance in predetermined characteristics for the first-time teaching sample in Appendix Table A13.

3.2 Professional performance

To evaluate the impact of the intervention on the professional performance of course participants, we employ publicly available data from the Judiciary of Peru's "Know Your Judge" (*Conoce a tu Juez*, in Spanish) platform.[11] This platform makes available information on several performance measures for individual Peruvian judges. We focus on judge-level measures referring to the professional performance in 2021—the year following our intervention—of judges who participated in levels 2 and 3 of the APP program in 2020. Since there are no data available on the professional performance of prosecutors, we restrict our analysis of professional outcomes to judges.

Performance measures available in the platform include the number of *verdicts* and the number of overall *decisions* per judge. Verdicts, in this context, constitute final decisions made by judges, based on the merit of the case. Meanwhile, the number of overall decisions consists of the number of cases decided for any reason by the judge; in addition to verdicts, it includes cases closed due to dismissal orders, final decrees, and other procedural reasons.

Also available through "Know Your Judge" is information on the judges' participation in a program titled "The Judge Listens to You" (*El Juez te Escucha*, in Spanish; henceforth, we refer to it as JLY). The program launched in October 2019 with the goal of bringing justice closer to the general public by fostering direct communication between litigants, their lawyers, and judges. Initially piloted in the civil and constitutional high courts of Lima, it was expanded to additional civil courts in the country in 2021. The program facilitates online meetings where participants can discuss procedural aspects of ongoing cases. Such discussions might address issues like sentencing delays, file assessments, or procedural flaws, without encroaching on substantive matters that are reserved for formal hearings to ensure all parties' rights are preserved.[12] The "Know Your Judge" platform contains, for each judge, information on the number of JLY meetings requested by litigants, as well as the number of meetings attended, canceled, or closed due to a no-show by the parties or the judge.

[11] The following website (in Spanish) provides further information on the platform: https://www.gob.pe/14476-consultar-informacion-sobre-magistrado-conoce-a-tu-juez. We extracted the data used in our analysis in September 2023.

[12] Per Administrative Resolution 000077-2021-CE-PJ, published in El Peruano.

To emphasize the importance of adhering to the JLY program, the platform explicitly lists the attendance of meeting requests as a performance metric of judges.

In addition to the data from the "Know Your Judge" platform, the Judiciary of Peru makes available case-level data with detailed information on court orders and other case-related events.[13] Using these data, we compute the number of cases filed in each judge's docket in 2021.[14] The case-level data are available from June to December 2021; all professional performance outcomes using these data are therefore measured over this period. The case-level data does not include information on family cases. We, therefore, exclude judges specialized in family cases from the analysis relying on these data. One concern is that judge performance as recorded in the KYJ platform could affect demand for judicial services, compromising the comparability of caseloads across treatment and control judges. In Peru, incoming cases are distributed across judges through an automated random assignment system, which substantially limits the scope for demand-side sorting in response to publicly visible performance metrics.[15] To assess this empirically, we compare total cases filed in 2021 across the two groups and find no significant difference (342 vs. 304 cases on average; $p = 0.41$). This result holds when regressing cases filed on treatment intensity, controlling for location and participant level ($p = 0.89$), suggesting caseloads are balanced across treatment and control groups.

With the available data on professional performance, we construct indicators to assess both the productivity of judges and the quality of case proceedings.[16] We define below the exact outcome variables that we use in our analysis, organized across these categories.

[13]For further information on these data (in Spanish), see https://www.gob.pe/14192-consultar-casilleros-digitales-de-jueces-y-juezas. We extracted the data used in our analysis in September 2023.

[14]Unfortunately, the case-level data do not contain clear information on case resolution. We are thus unable to compute measures such as case length; the direction of judges' decisions; or even the number of decisions and verdicts by a judge in a given year, which we would have been able to compare with the metrics from the "Know Your Judge" platform.

[15]Protocolo 06: Plan de Control Contra el Direccionamiento y Ruleteo de Demandas, Oficina de Control de la Magistratura (OCMA). Available at: https://anc.pj.gob.pe/contenido/documentos/protocolos/Protocolo_6.pdf.

[16]It is important to highlight that certain metrics initially proposed in the PAP, such as the case length, the fraction of decisions appealed or reversed, the length and direction of judicial decisions, or the textual measures of implicit bias, were not measurable due to data limitations (see footnote 14). Nevertheless, we were able to include in our analysis other valuable indicators—notably those derived from the JLY program, which was only implemented in large scale after our initial PAP submission in early 2020. Indicators related to this program provide relevant insight into the quality of judicial proceedings.

Productivity: We measure productivity by computing the judge's *case clearance rate*, which is the ratio of cases resolved (*number of overall decisions*, as reported in the "Know Your Judge" platform) to cases filed (obtained from the case-level data). An increase in the case clearance rate is associated with higher judicial productivity.

Quality of proceedings: As a first metric for quality, we compute the *rate of verdicts*, defined as the ratio between the *number of verdicts* and the *number of overall decisions* (both statistics are from the "Know Your Judge" platform).[17] We interpret a higher value as indicating a more in-depth investigation of the matter of the case, which leads to a final verdict. A lower value means that more cases ended due to procedural rather than substantive reasons.

More broadly, we view a higher verdict rate as desirable because it increases the likelihood that litigants receive a full adjudication of their claims. Decisions on the merits provide parties with a definitive resolution and a reasoned judgment addressing the substance of the dispute, which procedural terminations typically do not. By contrast, cases closed for procedural reasons often end without a substantive assessment of the underlying issues, potentially leaving claims only partially examined. At the same time, we acknowledge that verdicts are not always the socially optimal outcome: in some instances, early procedural closure may be efficient or even beneficial for one or both parties. Nonetheless, a verdict at least preserves the opportunity for the merits of the dispute to be considered and decided, whereas procedural closure forecloses that possibility altogether.

We also measure the quality of proceedings by considering the judges' participation in the JLY program. Specifically, we define, for each judge, the *attendance rate* as the ratio between the number of JLY audiences successfully completed and the number of audiences requested by the litigants. A higher value of this rate indicates more engagement of the judge with the public, which we interpret as better service provision by the judge. Additionally, we consider, as a separate metric, the number of *meetings requests*—that is, the number of JLY meeting requests from litigants that the judge received. A higher value of this metric indicates greater demand for judicial interaction. To be sure, the interpretation of this metric is not as straightforward as that of the attendance rate. A large number of meeting requests can suggest that the litigants see the judge as more accessible, which is aligned with the mission of the JLY program. Conversely, it can also reflect an increase in procedural issues that require

[17]The Justice Studies Center of the Americas (Centro de Estudios de Justicia de las Américas , CEJA) uses the same indicator—the ratio of verdicts to total decisions—as a proxy for whether cases are resolved on their merits rather than through procedural termination.

further discussion with the judge. In interpreting results related to the number of meeting requests below, we keep in mind this potential ambiguity.

Sample construction and data linkage. Since the professional performance outcomes draw on different data sources with different coverage, the estimation samples vary across specifications (Table 3). Of the 186 judges in the APP program, 181 were matched to the "Know Your Judge" platform. Of these, 173 have non-missing data on verdicts and decisions, yielding the sample for the verdict rate analysis (173 for the full sample and 87 for judges presiding on single-judge courts; see Section 4.2 for details on the latter subsample). The case clearance rate additionally requires data on cases filed, which we obtain from the separate case-level dataset; after excluding judges who specialize in family cases—for whom case filing data are unavailable—this yields 161 judges in the comprehensive sample and 84 in the single judge subsample. For the JLY outcomes, coverage is more limited: only 44 judges have JLY data, as the program had not yet been extended to all courts in 2021. To verify that these differences in sample composition do not introduce systematic selection, we present balance tests for each subsample in Appendix Table A12. These tests show no significant imbalance on observable characteristics across samples, including treatment exposure and caseload. In addition, we show in Appendix Figure A4 that the first-stage effects on grades are very similar across the judge samples used in Table 3, further reducing concerns that the professional-outcome results are driven by sample selection.

4 Results

This section first presents the regression results for educational outcomes (grades and satisfaction) of all students—that is, both judges and prosecutors. Then, we evaluate the impact of our intervention on professional outcomes, with a focus on judges; as explained in Section 3.2, data on prosecutors' professional performance are not available.[18] Finally, we assess the mechanisms behind our baseline results by focusing on two distinct components of the intervention—monitoring and feedback.

[18]We retain both prosecutors and judges in our analysis of the educational outcomes as the intervention operates through instructors and pedagogical practices that are common across judges and prosecutors within the same class. Including prosecutors allows us to more precisely estimate how the intervention affects classroom outcomes—including student performance, satisfaction, and instructor teaching behavior—and to better characterize how improvements in teaching translate into learning outcomes.

4.1 Grades and Satisfaction

To estimate the impact of the intervention on grades, we run the following regression specification:

$$Grades_{icr} = \alpha + \beta \times Treat_{cr} + \lambda \times \mathbf{Topic}_{cr} + \eta_r + \gamma \times \mathbf{Z}_{icr} + \varepsilon_{icr}, \quad (1)$$

where $Grades_{icr}$ is a grade score for student i in class c during round r; $Treat_{cr}$ is an indicator that class c received the intervention during round r; $\mathbf{Topic}_{cr}$ is a vector of dummies indicating the content of the course taught to class c in round r; η_r is a round fixed effect; and Z_{index} is a vector of control variables used in the stratification (namely, class location and participant level). The coefficient of interest, β, indicates the average impact of the intervention on grades. Standard errors are adjusted for clustering at the class level.

We proceed in a similar fashion to estimate the effects on satisfaction. Since there are two satisfaction surveys per round, we include the index s to account for the session. The regression specification is:

$$Satisfaction_{iscr} = \alpha + \beta \times Treat_{cr} + \lambda \times \mathbf{Topic}_{cr} + \eta_r + \theta \times Second_s + \gamma \times \mathbf{Z}_{imcr} + \varepsilon_{iscr}, \quad (2)$$

where $Satisfaction_{iscr}$ is a satisfaction measurement for student i in synchronous session s in class c and round r; and the remaining variables are as defined as in the specification for grades—with the addition of $Second_s$, which is a dummy indicating the second synchronous session in the round. As before, standard errors are adjusted for clustering at the class level.

Table 2 shows the main results on grades and satisfaction outcomes. We present the results for five evaluations taken by the students throughout the course (columns (1)-(5)) and the average satisfaction with the instructor and the course in the two class sessions (columns (6)-(7)). To facilitate the interpretation of the effects, we report estimates after standardizing each outcome based on the control group participants' mean and standard deviation.

First, we find an increase in final grades of approximately 0.22 standard deviations relative to the control group mean. The final grade is a weighted sum of all other course grades and, as a general assessment of performance in the course, is the most consequential grade for students. The intervention also produced significant increases in reading, homework, and exam grades, with magnitudes ranging from roughly 0.07

Table 2. 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
Treatment	0.1369	0.0732*	0.1747**	0.2366*	0.2196**	0.1443***	0.1348***
	(0.1063)	(0.0371)	(0.0805)	(0.1197)	(0.0803)	(0.0444)	(0.0482)
Control mean	17.68	18.91	17.96	17.38	17.73	0.7810	0.7700
Number of Participants	595	595	595	595	595	595	595
Observations	3,396	3,410	3,434	3,420	3,436	6,817	6,817
R^2	0.15663	0.17373	0.13176	0.08867	0.08940	0.02402	0.02694
Round fixed effects	✓	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: Standard errors in parentheses, clustered at the class level. All columns include strata controls (class level and location), as well as round and course fixed effects. Columns 6-7 also include session fixed effects. These regressions only include the first time a teacher teaches a course, which reduces the number of participants from 604 to 595. The number of participants is the unique number of students, whereas unit of observation is participant-round in columns 1-5 and participant-round-session in columns 6-7. Minor variations in number of observations are due to some missing outcome data points for very few participants. Results are robust to alternative specifications. Grades and satisfaction outcomes are standardized with respect to the control group mean. *p<0.10; **p<0.05; ***p<0.01.

to 0.24 standard deviations. The forum grade is the only component on which we do not detect a significant effect.

To assess whether these average effects reflect broad-based improvements or are driven by a subset of students, we examine the cumulative distribution functions of final grades and average satisfaction by treatment status. In both cases, the treated distribution lies to the right of the control distribution, indicating that the intervention improved outcomes across the distribution rather than only at the top or bottom (Figure A2).

Turning to satisfaction in columns (6) and (7), we find significant increases in satisfaction with both the instructor and the course, on the order of 0.14 standard deviations. Satisfaction is reported after every full day of classes, and the Academy requires students to complete these surveys, so attrition is not a concern for these outcomes. Appendix Table A1 reproduces Table 2 without the inclusion of control variables; the results are qualitatively and quantitatively similar.

Appendix Table A2 reports the same specifications on the full sample of class-round combinations—including rounds in which an instructor had previously taught a treated or control class earlier in the program. The main results hold in this specification; while the coefficients are slightly smaller in magnitude and less precisely estimated, the final grade and both satisfaction measures remain statistically significant. This pattern is consistent with potential spillovers across rounds: the broader sample blends the treatment effect on instructors' first program appearance with residual effects on instructors who later teach control classes, attenuating the average

effect.

A potential concern about the results in Table A2 is that the treatment status of the students in any round of our analysis is influenced by their class assignment in the first round of the program. To address this concern, Appendix Table A3 replicates the analysis in Table A2, but with the standard errors clustered at the level of the students' class assignment in the first round. Reassuringly, we obtain results that are qualitatively identical and quantitatively very similar to those in Table A2.

Appendix Table A4 presents the results for grades and satisfaction when the sample is split between judges and prosecutors. Despite the reduced sample size, which particularly affects the statistical power for judges, we still observe a significant increase in both grades and satisfaction for prosecutors, along with a notable improvement in reading grades for judges. A formal test of equality of coefficients across the two groups is insignificant for nearly all grade outcomes, with the only marginal exception being the forum grade ($p = 0.045$), higher for prosecutors under the first-time teaching specification, and the reading grade, higher for judges under the full sample specification ($p = 0.096$) (Appendix Table A5). The rest of grade and satisfaction outcomes are not statistically significant, suggesting that the apparent differences across professions may be driven by limited statistical power for the judge subsample. For satisfaction outcomes, the equality test is rejected ($p \approx 0.04$), with larger effects for prosecutors than for judges. A plausible explanation is a ceiling effect: in the control group, judges report higher baseline satisfaction than prosecutors do—0.84 versus 0.76 (on a 0-to-1 scale) for satisfaction with the instructor, and 0.84 versus 0.75 for satisfaction with the course—leaving less room for improvement.

Taken together, our results on educational outcomes indicate that the intervention improved both the objective performance of students in the class and their subjective assessment of the course. These findings are consistent with improved teaching by the instructors, learning by the students, and a better experience in the course overall.

4.2 Professional performance

We next ask whether the improvements in classroom outcomes translated into changes in judges' professional performance in the year following our intervention. Since our experimental design randomized treatment assignment at the class level, and judges who participated in the APP program take multiple classes over the year, the degree to which judges were exposed to the treatment may vary from one judge to another. To leverage the variation in treatment intensity, we define the variable

$Percentage\,Treated_i$ as the proportion of courses that received the intervention, out of the total of APP courses taken by judge i in 2020. We then consider the following specification to test whether the intervention had an impact on judicial performance:

$$y_i = \alpha + \beta_1 Percentage\,Treated_i + \gamma \times \mathbf{Z}_i + \varepsilon_i, \tag{3}$$

where y_i is the professional outcome indicator for judge i (which can refer to the ratio of verdicts, the judge's case clearance rate, the number of JLY meetings requested by litigants, and the attendance rate to those requests); and $\mathbf{Z}_i$ refers to the strata controls, which include location and participant level.

We note that judges may participate in panel decisions or in single-judge courts. Panels are generally composed of three judges deciding higher-instance court cases, while single-judge courts are presided over by one judge, typically at a lower instance court. We report results for two samples: a comprehensive sample, which includes all judges regardless of court type, and a restricted sample, which includes only judges who exclusively made single-judge decisions during 2021. This comparison is not intended to identify peer effects in judicial panels, which would require exogenous variation in panel composition. Rather, the single-judge subsample serves as a robustness check. By restricting to judges who made only individual decisions, we ensure that the estimated treatment effect reflects the judge's own behavior rather than contamination from the treatment status of co-panelists.

Columns (1) and (2) of Table 3 report estimates for the judge's case clearance rate—with column (1) corresponding to the comprehensive sample and column (2) to the single-judge sample. We interpret these estimates cautiously, as the numerator and denominator of the clearance rate originate from different data sources. The number of overall decisions—the numerator—comes from the "Know Your Judge" platform, while the number of cases filed—the denominator—comes from the case-level data, which does not include family cases. Although we exclude judges who specialize in family cases, the remaining judges may still handle some family cases, which would appear in the numerator but not in the denominator. Additionally, the case-level data covers only the period from June to December 2021, whereas the number of decisions refers to the entire year. This disconnect between numerator and denominator introduces measurement error and, as a byproduct, produces some observations with very few cases filed, which mechanically inflate the clearance rate.

To prevent these extreme observations from exerting disproportionate influence on the estimates, we exclude observations below the 2.5th or above the 97.5th percentile

of the clearance rate distribution. Even after trimming, the point estimates in columns (1) and (2) are not negligible in magnitude—especially in the single-judge sample—but they are statistically insignificant across both specifications.[19] Taken at face value, these results suggest that the intervention had no detectable effect on the overall productivity of judges.

Table 3. Impact on judges' professional performance

	Dependent variable:					
	Rate of case clearings		Rate of verdicts		JLY attend rate	JLY aud. requests
	(1)	(2)	(3)	(4)	(5)	(6)
Percent. Treat	-0.012	0.864	0.136**	0.189**	0.269**	39.684
	(0.756)	(1.587)	(0.054)	(0.091)	(0.128)	(35.767)
Control Group Mean	1.72	1.63	0.28	0.26	0.73	84.63
Observations	160	84	173	87	44	44
R^2	0.012	0.017	0.080	0.128	0.293	0.125
Adjusted R^2	-0.033	-0.074	0.041	0.050	0.179	-0.017

Notes: All dependent variables are measures for the period June-December 2021. Standard errors in parentheses are HC1 robust. Specifications (1)–(2) restrict to (a) judges with non-missing caseload data, (b) exclude outliers in the clearance rate — observations outside the p2.5–p97.5 range (cutoffs: 0.000 and 114.750) — and (c) exclude judges who specialize in family cases, for whom case filing data are unavailable. Specification (2) further restricts to single-judge panels. Figure A5 provides a visual summary of the sample construction and data linkage across all specifications. Specifications (1)–(4) control for location and participant level; (5)–(6) control only for location. Specifications (3)–(4) include all court types (including family courts). The *rate of case clearings* is the number of cases closed relative to all cases filed. The *rate of verdicts* refers to cases resolved with a verdict relative to the total cases resolved by the judge. The *JLY attendance rate* is the proportion of requested audiences in the "Judge Listens to You" program successfully completed by the judge. The *JLY audience requests* is the total number of "Judge Listens to You" audience requests received by the judge. Control group mean is the mean of the dependent variable among judges with *percent_treat* ≤ 0.5 in the regression sample. *p<0.10; **p<0.05; ***p<0.01.

We now turn our attention to the effect of the intervention on the quality of judicial proceedings. Columns (3) and (4) report estimates for the verdict rate — the share of a judge's decisions that constitute substantive rulings on the merits, as opposed to procedural closures. For both the comprehensive and single-court samples, we find that treated judges substantially increased their verdict rate: the estimated effect is 13.6 percentage points for the comprehensive sample and 18.9

[19] Appendix Table A10, columns (1) and (2), present estimates without this trimming; the null result holds. Columns (5) to (8) of the same table show results using an alternative definition of the clearance rate—the number of verdicts over the number of cases filed—with and without trimming. We report this alternative definition in the appendix for completeness; regardless of the specification, we find small, non-significant effects.

percentage points for single-court judges. Given a control-group mean of roughly 28 percent, these magnitudes imply an increase of over 40 percent relative to baseline. This compositional shift, holding overall productivity unchanged, is consistent with treated judges engaging more thoroughly with the merits of their cases rather than closing them on procedural grounds.

As a additional metric for the quality of proceedings, we consider the judges' participation in the JLY program, described in detail in Section 3.2.[20] In Table 3, column (5), the dependent variable consists of the judge's attendance rate to JLY audiences—that is, the number of completed audiences over the number of audiences that were requested by the litigants. The results indicate that the treated judges increased their attendance rate by 26.9 percentage points relative to a control-group mean of 73 percent, implying a near-complete attendance among treated judges in the year after our intervention.

Column (6) of Table 3 examines the number of JLY audience requests received by the judge. The estimated effect is positive but statistically insignificant, indicating that the increase in attendance is not driven by a reduction in requests. That the point estimate is positive suggests that, if anything, treated judges obtain more requests than non-treated ones.[21] [22]

To summarize, the professional outcome results paint a coherent picture. Treated judges resolved a substantially larger fraction of their caseload through substantive rulings on the merits, and attended a significantly higher share of litigant-requested meetings. Both findings are consistent with judges who are more deliberate and engaged in their work—attentive to the substance of cases and responsive to the parties before them. The absence of a productivity effect indicates that the quality gains came at no expense to throughput.

[20] As explained in Section 3.2, the JLY program had not yet been extended to all courts in 2021; in particular, it had not been implemented in criminal and specialized courts. As a consequence, the sample of judges in our analysis of JLY participation data is smaller than the sample that we employ to evaluate the impact of our intervention on other professional outcomes.

[21] Appendix Table A11 shows that, controlling for participant level, the effect on audience requests becomes positive and statistically significant. As discussed in Section 3.2, it is not straightforward to interpret an increase in the number of audience requests, by itself, as a desirable outcome of the intervention. But the findings in the Appendix, at the very least, corroborate our conclusion that the increase in the attendance rate in Table 3, column (5) is not due to fewer audience requests being made.

[22] As a further check, we compute a normalized measure of meeting requests defined as the ratio of JLY meeting requests to the judge's total caseload (cases filed during June-December 2021). The normalized coefficient is small and insignificant ($\beta = -1.986$, $p = 0.44$). The reduction in sample size from 44 to 39 judges reflects the availability of caseload data, and the loss of precision is partly attributable to noise introduced by the denominator.

4.3 Mechanism: Personalized feedback

Given the substantial effects of the intervention on academic and professional outcomes, in this section we ask what drives them. The intervention was designed to operate through its effect on instructor behavior. Monitoring and personalized feedback were intended to address gaps in the delivery of teaching and enhance instructor behavior, improving the organization, clarity, and engagement of teaching; better teaching, in turn, could raise student learning and satisfaction in the classroom and ultimately shape how judges conducted proceedings back at work. The remainder of this section presents evidence consistent with this chain, and then examines which component of the intervention—monitoring or personalized feedback—was the primary driver of the observed improvements.

To separately identify the contributions of monitoring and feedback, we leverage a timing asymmetry between the two components. The monitoring component—observation and assessment by the specialist—operates throughout both class sessions, since instructors knew from the outset they would be observed. The feedback component, by contrast, only takes effect after the first session: the post-session meeting in which the specialist and instructor reflect on strengths and areas for improvement, and the instructor commits to specific changes via the "Agreements and Commitments Form." This timing implies that, if feedback is the primary driver, the quality of teaching should improve in the second session relative to the first. If monitoring alone were the main mechanism, one might expect effects to appear already in the first session; if feedback played an important role, effects should become larger in the second session. Two types of outcomes satisfy the requirement of consistent measurement across both sessions: instructor performance grades (collected by the specialist, available for treated classes only) and student satisfaction (available for all classes).

Table 4 shows that specialist grades for treated instructors improved in the second session relative to the first across all dimensions—teaching skills, content mastery, and class structure. The improvement is approximately twice as large for class structure as for the other two components. Class structure encompasses how the instructor opens and motivates the session, the teaching strategies employed, student engagement and feedback, and the closing stage's emphasis on key takeaways. That these more malleable, organizational elements improve the most—while content mastery and intrinsic teaching skills are harder to move—would be consistent with the feedback component being the active ingredient, although we cannot rule out monitoring effects

that grow over time.

Table 5 reports the satisfaction results. Treatment effects on satisfaction are small in the first session—marginally significant for satisfaction with the instructor and insignificant for satisfaction with the course (columns (1) and (4)). They become substantially larger and statistically significant in the second session (columns (2) and (5)). The session-2 minus session-1 difference is approximately 0.12 standard deviations for both satisfaction with the instructor and satisfaction with the course(columns (3) and column (6))—a pattern consistent with teaching quality improving more in the second session. Within the treatment group, the change in student satisfaction between sessions is significantly associated with the improvement in class structure and content mastery grades (Appendix Table A8), suggesting that gains in session organization and substantive clarity contributed to the observed satisfaction improvements, while teaching skills show no significant association.

A natural concern is whether these associations reflect grade inflation rather than genuine teaching improvements. Treated instructors may have graded participants more generously, and such generosity—rather than better instruction—could explain higher satisfaction scores. We address this concern in two ways. First, students complete satisfaction surveys immediately after each class session, before the final exam, so exam grades cannot mechanically influence satisfaction responses. Second, while intermediate grades (homework, readings, participation) are awarded during the course and could in principle be known to students at the time of the survey, replicating the satisfaction regressions while controlling for these grades leaves treatment effects unchanged (Appendix Table A9). Taken together, the evidence suggests that grade generosity is unlikely to drive the satisfaction results.

Collectively, these results suggest that the intervention was particularly effective at improving the quality of the teaching in the second session of the class. While our findings are consistent with the feedback component driving the effects of the intervention, we take them only as suggestive evidence, as we cannot rule out other potential mechanisms that could be behind the effect.[23] Additionally, we find that the improvement in class structure appears to be an important driver of the increase in student satisfaction and instructor performance. This result suggests that the intervention's success was due, at least in part, to its effect on more malleable aspects of instruction, such as class organization, motivational techniques, teaching strategies,

[23]For example, a competing hypothesis is that the monitoring component has a compounded effect, and becomes more effective with time—i.e., in the second session relative to the first one. While we consider this hypothesis less plausible, we cannot completely rule it out based on the available data.

student engagement, and the closing and review stages.

While the measured improvement in instructor scores may appear modest in absolute terms, the intervention operates in a setting where even incremental changes in teaching quality can affect a large number of judicial decisions. Training is delivered over an eight-month course, allowing improvements in instruction to accumulate throughout the learning process. Moreover, each trained judge subsequently handles a high volume of cases, so gains in learning and workplace practice can translate into meaningful cumulative effects on observed outcomes.

Table 4. Instructor Grades per Session

Dependent Variable:	Teaching Skills	Content Mastery	Class structure
Model:	(1)	(2)	(3)
Variables			
session2	0.4144***	0.4417***	0.8308***
	(0.1252)	(0.1255)	(0.1393)
Fixed-effects			
Course	Yes	Yes	Yes
Round	Yes	Yes	Yes
Class	Yes	Yes	Yes
Fit statistics			
Observations	132	132	132
R^2	0.56897	0.53287	0.64139
Within R^2	0.07019	0.07815	0.24506

Notes: Sample restricted to treated classes only and to the first round each instructor taught, as specialist evaluation scores are only available for treated instructors. Each column represents a different outcome, computed as the average of all grading questions within that category; each item is scored on a 0–20 scale. All regressions include course, round, and class fixed effects. Standard errors in parentheses are clustered at the course level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Conclusion

The performance of judges and prosecutors is essential for a well-functioning justice system—a core component of state capacity. This is particularly important in developing countries such as Peru, where the justice system is seen as opaque and lacks citizens' trust. In this study, we partnered with the Judicial Academy of Peru (AMAG) to implement an RCT within the Academy's Promotion Program (APP), a program that trains active judges and prosecutors seeking promotion every year. We

Table 5. Personalized feedback effects on satisfaction

	Satisfaction with teacher			Satisfaction with course		
	(1) First session	(2) Second session	(3) Diff.	(4) First session	(5) Second session	(6) Diff.
Treatment	0.0811*	0.2078***	0.1228***	0.0722	0.1977***	0.1213***
	(0.0445)	(0.0481)	(0.0262)	(0.0477)	(0.0515)	(0.0231)
Number of Participants	595	595	595	595	595	595
Observations	3,411	3,406	3,405	3,411	3,406	3,405
R^2	0.02384	0.02948	0.01454	0.02891	0.03068	0.01475
Round fixed effects	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓

Note: Sample restricted to the first round each instructor taught in the program. Columns (1) and (4) report estimates for the first session subsample; columns (2) and (5) for the second session subsample; columns (3) and (6) for the within-student difference between the two sessions. All columns include strata controls (location and participant level), as well as round and course fixed effects. Standard errors in parentheses are clustered at the class level. Satisfaction outcomes are standardized with respect to the control group mean. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

randomized whether instructors received an online monitoring and feedback program that aimed to improve the quality and effectiveness of their teaching.

We find that the intervention increased both class and job performance indicators. For judges and prosecutors in the treatment group, we observed higher grades and greater satisfaction with both the course and the instructor. Furthermore, we find significant effects on the quality of judicial proceedings conducted by judges in the treatment group in the year after the intervention. In specific, we find an increase in the ratio of verdicts to any decision, which is also accompanied by an increase in the rate at which judges attended meetings requested by litigants. Our analysis suggests that the feedback provided to instructors played a significant role in driving these improvements in teaching quality and student learning.

It is worth noting that our intervention took place in 2020, when AMAG transitioned abruptly to online delivery in response to COVID-19, and the program coincided with a broader curricular shift toward case-based teaching. These features of the setting are worth bearing in mind when considering generalizability. The shift to virtual formats and the curricular transition both increased the returns to the dimensions the intervention targeted — session organization, student engagement, and effective use of real-world examples — and the direction of any resulting bias on our

estimates is difficult to sign. While the pandemic represents an unusual episode, the broader features it exemplifies — abrupt transitions to new instructional formats, instructors without structured pedagogical support, and institutions undergoing curricular reform — are not uncommon in developing country settings. We therefore view these contextual factors as shaping the specific episode we study without fundamentally limiting the broader relevance of our findings.

To our knowledge, this is the first study to utilize a randomized experiment in the context of judicial training. The findings demonstrate the potential for enhancing the quality of education for judges and prosecutors, highlighting the positive impact that such improvements can have on the services provided to citizens. Given the substantial investments in judicial training by national and international organizations worldwide (USAID, World Bank), our results warrant further investigation in future research.

Additionally, our study contributes to the broader literature on state effectiveness and bureaucratic performance by showing how enhancing the quality of instruction for civil servants can improve the provision of public services. In contrast to prior research that evaluates newly-designed training modules (Azulai et al., 2020; Banerjee et al., 2021; Mehmood et al., 2024), we focus on enhancing the quality of the existing training of civil servants. Our results suggest that this approach, which has the potential to be applicable across various sectors beyond justice, leads to improved educational and professional outcomes for civil servants. Consequently, our research points to a promising new avenue for improving both the quality of instruction of public employees and public service delivery.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process. During the preparation of this work the authors used Claude to help with manuscript preparation and statistical analysis. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Acemoglu, Daron and Simon Johnson, "Unbundling institutions," *Journal of political Economy*, 2005, *113* (5), 949–995.

Allen, Joseph P, Robert C Pianta, Anne Gregory, Amori Yee Mikami, and Janetta Lun, "An interaction-based approach to enhancing secondary school instruction and student achievement," *Science*, 2011, *333* (6045), 1034–1037.

Amirapu, Amrit, "Justice delayed is growth denied: The effect of slow courts on relationship-specific industries in India," *Economic Development and Cultural Change*, 2021, *70* (1), 415–451.

Angrist, Noam, David K Evans, Deon Filmer, Rachel Glennerster, Halsey Rogers, and Shwetlena Sabarwal, "How to improve education outcomes most efficiently? A review of the evidence using a unified metric," *Journal of Development Economics*, 2024, *172*, 103382.

Araujo, M Caridad, Pedro Carneiro, Yyannú Cruz-Aguayo, and Norbert Schady, "Teacher quality and learning outcomes in kindergarten," *The Quarterly Journal of Economics*, 2016, *131* (3), 1415–1453.

Ash, Elliott, Daniel L Chen, and Suresh Naidu, "Ideas have consequences: The impact of law and economics on american justice," Technical Report, National Bureau of Economic Research 2022.

Azulai, Michel, Imran Rasul, Daniel Rogger, and Martin J Williams, "Can training improve organizational culture? Experimental evidence from Ghana's civil service," Technical Report, EDI Working Paper Series 2020.

Banerjee, Abhijit, Raghabendra Chattopadhyay, Esther Duflo, Daniel Keniston, and Nina Singh, "Improving police performance in Rajasthan, India: Experimental evidence on incentives, managerial autonomy, and training," *American Economic Journal: Economic Policy*, 2021, *13* (1), 36–66.

Baye, Michael R and Joshua D Wright, "Is antitrust too complicated for generalist judges? The impact of economic complexity and judicial training on appeals," *The Journal of Law and Economics*, 2011, *54* (1), 1–24.

Best, Michael Carlos, Jonas Hjort, and David Szakonyi, "Individuals and organizations as sources of state effectiveness," *American Economic Review*, 2023, *113* (8), 2121–2167.

Centro de Estudios de Justicia de las Américas (CEJA), "Los indicadores judiciales: herramienta para la aplicación de la gestión por desempeño," Technical Report, Centro de Estudios de Justicia de las Américas, Santiago, Chile 2008.

Chakraborty, Pavel, "Judicial quality and regional firm performance: The case of Indian states," *Journal of Comparative Economics*, 2016, *44* (4), 902–918.

Chemin, Matthieu, "Does court speed shape economic activity? Evidence from a court reform in India," *The Journal of Law, Economics, & Organization*, 2012, *28* (3), 460–485.

_ , Daniel L Chen, Vincenzo Di Maro, Paul Kieti Kimalu, Momanyi Mokaya, and Manuel Ramos-Maqueda, "Data Science for Justice: The Short-Term Effects of a Randomized Judicial Reform in Kenya," Technical Report, Toulouse School of Economics Working Paper 2022.

Credential Engine, "Education and Training Expenditures in the U.S.," Technical Report, Washington, DC: Credential Engine 2021.

Cuddy, Colleen, "Cultivating communities of practice: A guide to managing knowledge," *The Bottom Line*, 2002, *15* (2).

Evans, David K and Anna Popova, "What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews," *The World Bank Research Observer*, 2016, *31* (2), 242–270.

Finan, Frederico, Benjamin A Olken, and Rohini Pande, "The personnel economics of the developing state," in "Handbook of economic field experiments," Vol. 2, Elsevier, 2017, pp. 467–514.

Glewwe, Paul and Karthik Muralidharan, "Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications," in "Handbook of the Economics of Education," Vol. 5, Elsevier, 2016, pp. 653–743.

González, Leonel and Jeremy Cooper, *Judicial Training in Latin America: A Study on the Practices of Judicial Colleges*, Santiago, Chile: Justice Studies Center of the Americas (CEJA), 2017.

Hammergren, Linn, "Judicial Training and Justice Reform," Technical Report, U.S. Agency for International Development 1998.

Hansen, Jesper Asring and Lars Tummers, "A systematic review of field experiments in public administration," *Public Administration Review*, 2020, *80* (6), 921–931.

Kondylis, Florence and Mattea Stein, "The speed of justice," *Review of Economics and Statistics*, 2023, *105* (3), 596–613.

Kremer, Michael, Conner Brannen, and Rachel Glennerster, "The challenge of education and learning in the developing world," *Science*, 2013, *340* (6130), 297–300.

Laeven, Luc and Christopher Woodruff, "The quality of the legal system, firm ownership, and firm size," *The Review of Economics and Statistics*, 2007, *89* (4), 601–614.

Mehmood, Sultan, Shaheen Naseer, and Daniel L Chen, "Altruism in governance: Insights from randomized training for Pakistan's junior ministers," *Journal of Development Economics*, 2024, p. 103317.

Osorio, Jorgeluis Alan Romero, "Reforma de la Academia de la Magistratura del Perú," *Revista Oficial del Poder Judicial*, 2020, *11* (13), 343–361.

Piper, Benjamin, Joseph Destefano, Esther M Kinyanjui, and Salome Ong'ele, "Scaling up successfully: Lessons from Kenya's Tusome national literacy program," *Journal of Educational Change*, 2018, *19*, 293–321.

Ramos-Maqueda, Manuel and Daniel Chen, "The Role of Justice in Development: The Data Revolution," Technical Report, The World Bank 2021.

Sadka, Joyce, Enrique Seira, and Christopher Woodruff, "Information and Bargaining through Agents: Experimental Evidence from Mexico's Labour Courts," *The Review of Economic Studies*, 2024, *91* (46), 3677–3711.

Schaffner, Julie, Paul Glewwe, and Uttam Sharma, "Why Programs Fail: Lessons for Improving Public Service Quality from a Mixed-Methods Evaluation of an Unsuccessful Teacher Training Program in Nepal," *The World Bank Economic Review*, 2024, *39* (2), 473–496. Published online July 8, 2024; journal volume dated May 2025.

The World Bank, "New Directions in Justice Reform: A Companion Piece to the Updated Strategy and Implementation Plan on Strengthening Governance, Tackling Corruption," Technical Report, The World Bank 2012.

Wenger, Etienne, "Communities of practice: Learning as a social system," *Systems thinker*, 1998, *9* (5), 2–3.

_ , "Communities of practice: A brief introduction," https://scholarsbank.uoregon.edu/items/6baf09ba-c319-4949-acef-cb220dc2953a 2011. Online; accessed December 12, 2024.

World Justice Project, "The Rule of Law in Peru: Key Findings from the General Population Poll 2022.," Technical Report, World Justice Project 2022.

A Appendix: Supplementary figures and tables

This appendix provides additional details on the intervention, data construction, and robustness checks. It includes supplementary figures and tables on treatment exposure, classroom outcomes, professional performance outcomes, sample construction, and mechanism checks.

- Figure A1: Distribution of participants by treatment intensity.
- Table A1: Treatment effects on grades and satisfaction (uncontrolled specifications).
- Table A2: Treatment effects on grades and satisfaction (full sample).
- Table A3: Treatment effects on grades and satisfaction, with standard errors clustered at the first-round class level.
- Table A4: Treatment effects on grades and satisfaction by profession (first-time teaching sample).
- Table A5: Treatment effects on grades and satisfaction by profession (full sample).
- Figure A2: Cumulative distribution of main grades by treatment status.
- Figure A3: Cumulative distribution of average satisfaction by treatment status
- Table A8: Association between student satisfaction and instructor grades.
- Table A9: Treatment effects on student satisfaction controlling for grades.
- Table A10: Impact on judges' clearance rate under alternative definitions.
- Table A11: Treatment effects on meetings with litigants (robustness checks).
- Figure A4: Running main grade specifications on different samples of judges
- Figure A5: Sample construction and data linkage across professional outcome specifications.

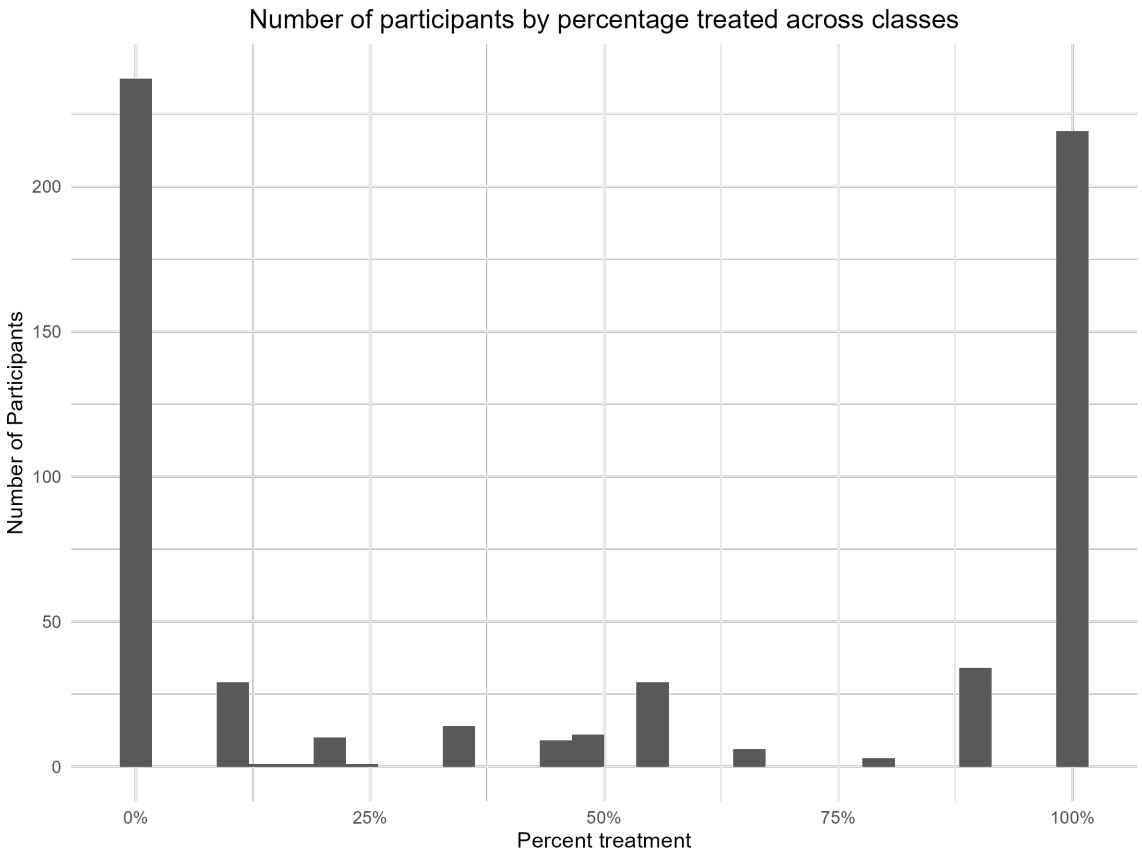

Figure A1. Participants by treatment intensity

Table A1. Treatment effects on grades and satisfaction (uncontrolled specifications)

	Grades					Satisfaction	
	(1) Forum	(2) Reading	(3) Homework	(4) Exam	(5) Final grade	(6) With teacher	(7) With course
Treatment	0.1439	0.0726	0.1350	0.2353*	0.2128**	0.1516**	0.1498**
	(0.1005)	(0.0520)	(0.0852)	(0.1258)	(0.0895)	(0.0571)	(0.0594)
Control mean	17.68	18.91	17.96	17.38	17.73	0.7810	0.7700
Number of Participants	595	595	595	595	595	595	595
Observations	3,396	3,410	3,434	3,420	3,436	6,817	6,817
R^2	0.00553	0.00128	0.00558	0.01346	0.01300	0.00607	0.00605

Notes: Standard errors in parentheses, clustered at the class level. No strata controls or fixed effects are included. Sample is restricted to the first time a teacher teaches a course, which reduces the number of participants from 604 to 595. Unit of observation is participant-round in columns 1-5 and participant-round-meeting in columns 6-7. Grades and satisfaction outcomes are standardized with respect to the control group mean. $*$ p $<$ 0.10, $**$ p $<$ 0.05, $***$ p $<$ 0.01

Table A2. Treatment effects on grades and satisfaction (full sample)

	Grades					Satisfaction	
	(1) Forum	(2) Reading	(3) Homework	(4) Exam	(5) Final	(6) Teacher	(7) Course
Treatment	0.0677	0.0568	0.0882	0.1584	0.1257**	0.0997*	0.1013*
	(0.0764)	(0.0353)	(0.0539)	(0.0964)	(0.0605)	(0.0545)	(0.0526)
Control mean	17.82	19.05	17.98	17.54	17.87	0.7860	0.7740
Number of Participants	604	604	604	604	604	604	604
Observations	5,033	5,053	5,090	5,067	5,094	10,099	10,099
R^2	0.13759	0.15011	0.11455	0.06350	0.08060	0.02614	0.02967
Round fixed effects	✓	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: Standard errors in parentheses are clustered at the class level. All columns include strata controls (location and participant level), as well as round and course fixed effects. Columns (6)–(7) also include session fixed effects. The unit of observation is participant-round in columns (1)–(5) and participant-round-session in columns (6)–(7); the number of participants refers to unique students. Minor variations in observation counts across columns reflect missing outcome data for a small number of participants. Grades and satisfaction outcomes are standardized with respect to the control group mean; grades are on a 0–20 scale and satisfaction on a 0–1 scale. *p<0.10; **p<0.05; ***p<0.01.

Table A3. Treatment effects on grades and satisfaction, clustered at 1st round's class (full sample)

	Grades					Satisfaction	
	(1) Forum grade	(2) Reading grade	(3) Homework grade	(4) Exam grade	(5) Final grade	(6) With teacher	(7) With course
Treatment	0.0686	0.0579	0.0891	0.1589	0.1263*	0.0993*	0.1004*
	(0.0777)	(0.0385)	(0.0542)	(0.0926)	(0.0648)	(0.0553)	(0.0574)
Number of Participants	604	604	604	604	604	604	604
Observations	5,033	5,053	5,090	5,067	5,094	10,099	10,099
R^2	0.13792	0.15050	0.11485	0.06360	0.08072	0.02618	0.02990
Round fixed effects	✓	✓	✓	✓	✓	✓	✓
Course fixed effects	✓	✓	✓	✓	✓	✓	✓

Notes: Standard errors in parentheses, clustered using the class they were assigned in the first round. All columns include strata controls (class level and location), as well as round and course fixed effects. Columns 6-7 also include meeting fixed effects.The number of participants is the unique number of students, whereas unit of observation is participant-round in columns 1-5 and participant-round-meeting in columns 6-7. Minor variations in number of observations are due to some missing outcome data points for very few participants. Standard errors are clustered at the class level. Results are robust to alternative specifications. Grades and satisfaction outcomes are standardized with respect to the control group mean. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$

Table A4. Treatment effects on grades and satisfaction by profession

	Grades					Satisfaction	
	(1) Forum	(2) Reading	(3) Homework	(4) Exam	(5) Final	(6) With teacher	(7) With course
Panel A: Prosecutors							
Treatment	0.1907*	0.0337	0.2171*	0.2695*	0.2595**	0.2098***	0.1853***
	(0.1096)	(0.0488)	(0.1068)	(0.1325)	(0.1087)	(0.0540)	(0.0534)
Participants	404	404	404	404	404	404	404
Observations	2326	2339	2355	2347	2357	4683	4683
R Squared	0.154	0.179	0.140	0.093	0.105	0.027	0.032
Panel B: Judges							
Treatment	0.0136	0.1595***	0.0635	0.1629	0.1340	-0.0358	-0.0122
	(0.1136)	(0.0396)	(0.0845)	(0.1184)	(0.0877)	(0.0686)	(0.0700)
Participants	181	181	181	181	181	181	181
Observations	1059	1060	1068	1062	1068	2112	2112
R Squared	0.192	0.181	0.146	0.102	0.083	0.054	0.044
p-value (Prosecutor = Judge)	0.045	0.143	0.274	0.154	0.327	0.040	0.043

Note: Standard errors are clustered at the class level. Grades and satisfaction outcomes are standardized with respect to the control group mean. All regressions control for location (randomization strata), participant baseline level, and include round and course fixed effects. Columns 6-7 also include meeting fixed effects. Regressions are estimated on the subsample of the first time a teacher teaches a given course. Panel A shows regression coefficients for the prosecutor subsample. Panel B shows regression coefficients for the judge subsample. The row "p-value (Prosecutor = Judge)" reports the p-value from a test of equality of the treatment coefficients across the two subsamples, obtained from a pooled regression interacting treatment with an indicator for prosecutors. *p<0.10; **p<0.05; ***p<0.01.

Table A5. Treatment effects on grades and satisfaction by profession (full sample)

	Grades					Satisfaction	
	(1) Forum	(2) Reading	(3) Homework	(4) Exam	(5) Final	(6) With teacher	(7) With course
Panel A: Prosecutors							
Treatment	0.1020	0.0217	0.1273*	0.1904*	0.1529*	0.1727***	0.1565**
	(0.0808)	(0.0436)	(0.0735)	(0.1079)	(0.0873)	(0.0600)	(0.0569)
Participants	407	407	407	407	407	407	407
Observations	3439	3458	3482	3467	3486	6922	6922
R Squared	0.124	0.155	0.118	0.070	0.092	0.034	0.037
Panel B: Judges							
Treatment	-0.0098	0.1481***	-0.0187	0.0881	0.0586	-0.1068	-0.0619
	(0.0854)	(0.0456)	(0.0860)	(0.1007)	(0.0801)	(0.0756)	(0.0703)
Participants	183	183	183	183	183	183	183
Observations	1578	1579	1592	1584	1592	3147	3147
R Squared	0.173	0.157	0.145	0.072	0.087	0.041	0.034
p-value (Prosecutor = Judge)	0.154	0.096	0.326	0.189	0.500	0.006	0.007

Notes: 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 prosecutor subsample. Panel B shows regression coefficients for the judge subsample. The row "p-value (Prosecutor = Judge)" reports the p-value from a test of equality of the treatment coefficients across the two subsamples, obtained from a pooled regression interacting treatment with an indicator for prosecutors. *p<0.10; **p<0.05; ***p<0.01.

Table A6. Sample composition by course topic and treatment status

Round	Course	N Classes		N participants	% Treated
		Control	Treated		
1	Fundamentals of Ethics	2	2	126	50.0
1	Legal Argumentation	9	9	452	50.2
2	Constitutional Theory	2	2	128	50.0
2	Ethics in the Judiciary	9	8	451	47.2
3	Conventionality and Constitutionality (Basic)	2	2	131	49.6
3	Conventionality and Constitutionality (Advanced)	9	9	451	50.3
4	Legal Interpretation	2	2	129	50.4
4	Constitutional Procedure: Cases	9	9	451	50.1
5	Criminal Case Methodology	2	2	108	51.9
5	Constitution and Administrative Law	9	9	450	50.2
6	Problems in Civil Process	1	0#	25	0.0
6	Criminal and Criminal Procedure Debates	2	2	103	51.5
6	Evidence on Appeal (Civil)	1	2	65	66.2
6	Evidence on Appeal (Criminal)	9	7	384	43.8
7	Criminal Appeals Law	1	0#	26	0.0
7	Supreme Court Criminal Jurisprudence	2	2	104	51.9
7	Civil Appeals Law	1	2	66	66.7
7	Supreme Court Criminal Jurisprudence	9	7	382	44.2
8	Family Law: Cases	1	0#	29	0.0
8	Forensic Investigation	2	2	98	51.0
8	Supreme Court Civil Agreements	1	2	62	66.1
8	Supreme Court Criminal Agreements	8	7	383	48.3
9	Judicial Office Management	2	2	100	49.0
9	Tax Office Management	1	0#	26	0.0
9	Public Management Skills	7	8	447	53.7

*Notes:*Each row corresponds to one course in one round. Class columns show the number of classes assigned to each treatment arm. N participants is the total number of students enrolled in the course. % Treated is the percentage of enrolled participants in the treated classes.

# denotes courses with zero classes in one or both treatment arms — these are specialized courses in rounds 6–9, absorbed by *course* fixed effects.

Table A7. Instructor assignment to treatment and control classes, by rounds taught

Rounds taught	Instructors N	treatment only	control only	both
1	85	38	47	0
2	34	9	6	19
3	13	3	1	9
4	2	1	0	1
Total	**134**	**51**	**54**	**29**

Note: This table summarises the assignment of instructors to treatment and control classes across all 9 rounds of the APP program. An instructor is classified as 'treatment only' if they taught exclusively in treated classes, 'control only' if they taught exclusively in control classes, and 'both' if they taught in both treated and control classes across different rounds.

Table A8. Association between Student Satisfaction and Instructor Grades

	Content Mastery		Class Structure		Teaching Skills	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.2648*	0.2824**	0.6241***	0.6346***	0.3754***	0.3561***
	(0.1401)	(0.1290)	(0.1544)	(0.1413)	(0.1393)	(0.1287)
Average Satisfaction	3.671*		4.290**		0.8100	
	(1.852)		(2.040)		(1.841)	
Satisfaction with the teacher		3.985**		4.912**		1.458
		(1.795)		(1.966)		(1.791)
Observations	66	66	66	66	66	66
R^2	0.05783	0.07151	0.06461	0.08890	0.00301	0.01026
Adjusted R^2	0.04311	0.05700	0.04999	0.07466	-0.01256	-0.00521

Notes: This table shows the results of a linear regression of the change in satisfaction between the second and the first session on the change in instructor grading between the second and the first session. "Average satisfaction" refers to the overall average of all responses to the satisfaction question, whereas "Satisfaction with the teacher" refers to the average of questions related only to satisfaction with the teacher. The different dependent variables refer to different dimensions that the instructor received grades on. The regressions shown do not include any controls. Sample is restricted to the first time a teacher teaches a course. Results remain similar if fixed effects on class or round are included. SE shown in parantheses are HC1 robust. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$

Table A9. Treatment effects on satisfaction, controlling for grades

	Satisfaction with teacher		Satisfaction with course	
	(1)	(2)	(3)	(4)
Treatment	0.1524***	0.1594***	0.1419***	0.1469***
	(0.0449)	(0.0452)	(0.0478)	(0.0475)
Control mean	0.7810	0.7810	0.7700	0.7700
Intermediate grades	Yes	Yes	Yes	Yes
Exam grade	No	Yes	No	Yes
Observations	6,770	6,766	6,770	6,766
R^2	0.02672	0.02739	0.02968	0.03002
Within R^2	0.01488	0.01553	0.01316	0.01334
round fixed effects	✓	✓	✓	✓
idcurso fixed effects	✓	✓	✓	✓

Notes: Standard errors clustered at the class level. All specifications include round and course fixed effects, location strata, and participant level controls. Intermediate grades include forum, reading, and homework grades. All-grades specification additionally controls for the exam grade. Grades are standardized with respect to the control group mean. Sample is restricted to the first time a teacher teaches a course. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$

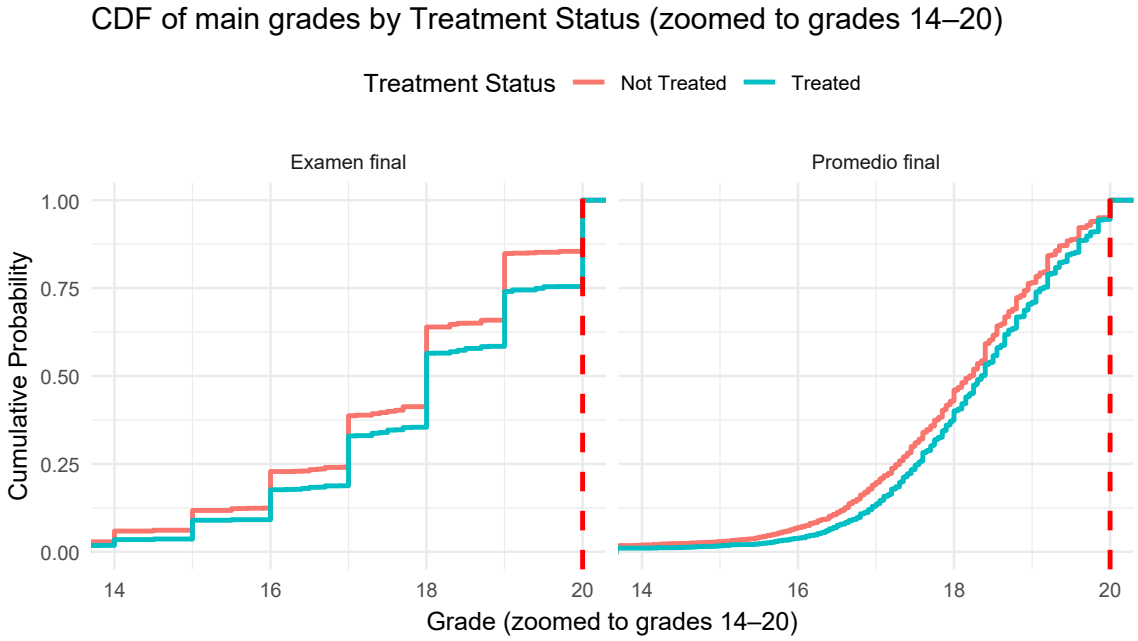

Figure A2. CDF of main grades by treatment status.

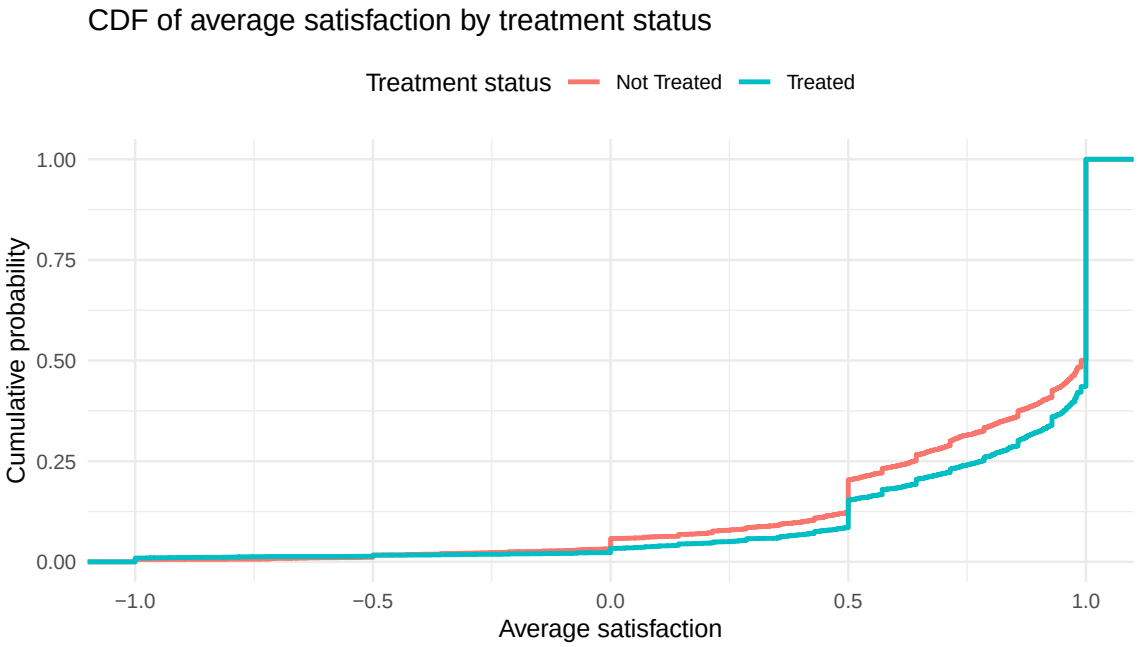

Figure A3. CDF of average satisfaction by treatment status.

Table A10. Impact on judges' clearance rates using alternative output measures

	Dependent variable: Rate of case clearings							
	All decisions / cases filed				Substantive verdicts / cases filed			
	All	Single	All	Single	All	Single	All	Single
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Percent. Treat	3.633	0.864	-0.012	0.864	0.231	0.320	0.018	-0.155
	(3.776)	(1.587)	(0.756)	(1.587)	(0.292)	(0.521)	(0.198)	(0.188)
Control Group Mean	1.72	1.63	1.72	1.63	0.38	0.32	0.38	0.32
Observations	161	84	160	84	161	84	160	83
R^2	0.015	0.017	0.012	0.017	0.022	0.023	0.015	0.036
Adjusted R^2	-0.030	-0.074	-0.033	-0.074	-0.023	-0.067	-0.030	-0.054
Single judges only		✓		✓		✓		✓
Outliers trimmed			✓	✓			✓	✓

Notes: Standard errors in parentheses are HC1 robust. All specifications restrict to judges with non-missing caseload data and exclude family courts; observations with missing court type are retained. Columns (1)–(4) define the clearance rate as all decisions (*sentencias* plus *sentidos de fallo*) divided by total cases filed. Columns (5)–(8) define the clearance rate as substantive verdicts (*sentencias*) divided by total cases filed. Columns (2), (4), (6), and (8) restrict the sample to judges who decide cases as single judges. Columns (3)–(4) exclude outliers in the all-decisions clearance rate; columns (7)–(8) exclude outliers in the substantive-verdict clearance rate. Outliers are observations outside the p2.5–p97.5 range among judges with non-missing caseload data. All dependent variables are measured over June–December 2021. All specifications control for location and participant level. Control group mean is the mean of the dependent variable among judges with *percent_monit* ≤ 0.5 in the regression sample. *p<0.10; **p<0.05; ***p<0.01.

Table A11. Impact of Treatment on Judges' Attendance Rates

	Total Meetings	Attendance Rate
	(1)	(2)
Percentage Monitored	85.129*	0.305*
	(47.697)	(0.179)
participant_level_2	60.872	0.045
	(58.481)	(0.134)
Dep Var Mean	53	0.68
N	44	44
R^2	0.176	0.298
Adjusted R^2	−0.013	0.137

Notes: HC1-robust Standard Errors in parantheses. The regression controls for location and participant level strata. Total meetings refer to the number of meetings a judge scheduled during 2021. Attendance rate refers to the rate of meetings that the judge and litigant both attended during this same period. The number of judges in this regression is 44, as these are the sample of judges who participated in the pilot program of "El Juez te Escucha" out of the total sample of 193 judges.$*$ p $<$ 0.10, $**$ p $<$ 0.05, $***$ p $<$ 0.01

Table A12. Balance on Observables by Table 3 Regression Sample

Characteristic	Beta	(Robust SE)	p-value	N total	N control	N treated
Panel A: Clearance rate — all judges						
Female	-0.078	(0.079)	0.325	161	86	75
Age	-0.495	(0.950)	0.603	161	86	75
Years of tenure	-0.653	(0.528)	0.218	161	86	75
Years in bar association	-0.920	(0.800)	0.252	161	86	75
Panel B: Clearance rate — single judge						
Female	-0.085	(0.116)	0.466	84	48	36
Age	0.029	(1.412)	0.983	84	48	36
Years of tenure	0.231	(0.386)	0.551	84	48	36
Years in bar association	0.621	(1.094)	0.572	84	48	36
Panel C: Verdict rate — all judges						
Female	-0.079	(0.077)	0.306	173	92	81
Age	-0.785	(0.941)	0.405	173	92	81
Years of tenure	-0.465	(0.544)	0.394	173	92	81
Years in bar association	-1.125	(0.806)	0.165	173	92	81
Panel D: Verdict rate — single judge						
Female	-0.151	(0.114)	0.192	87	51	36
Age	-0.979	(1.419)	0.492	87	51	36
Years of tenure	0.239	(0.366)	0.516	87	51	36
Years in bar association	-0.601	(1.142)	0.601	87	51	36
Panel E: JLY Meetings						
Female	0.123	(0.163)	0.453	44	27	17
Age	-1.000	(2.651)	0.708	44	27	17
Years of tenure	1.030	(1.375)	0.459	44	27	17
Years in bar association	-0.502	(1.901)	0.793	44	27	17

Note: Each row reports the coefficient on the treatment indicator from an OLS regression of the balance variable on treatment with strata controls. Standard errors are HC1 robust. No outlier trimming applied. *** p<0.01, ** p<0.05, * p<0.10.

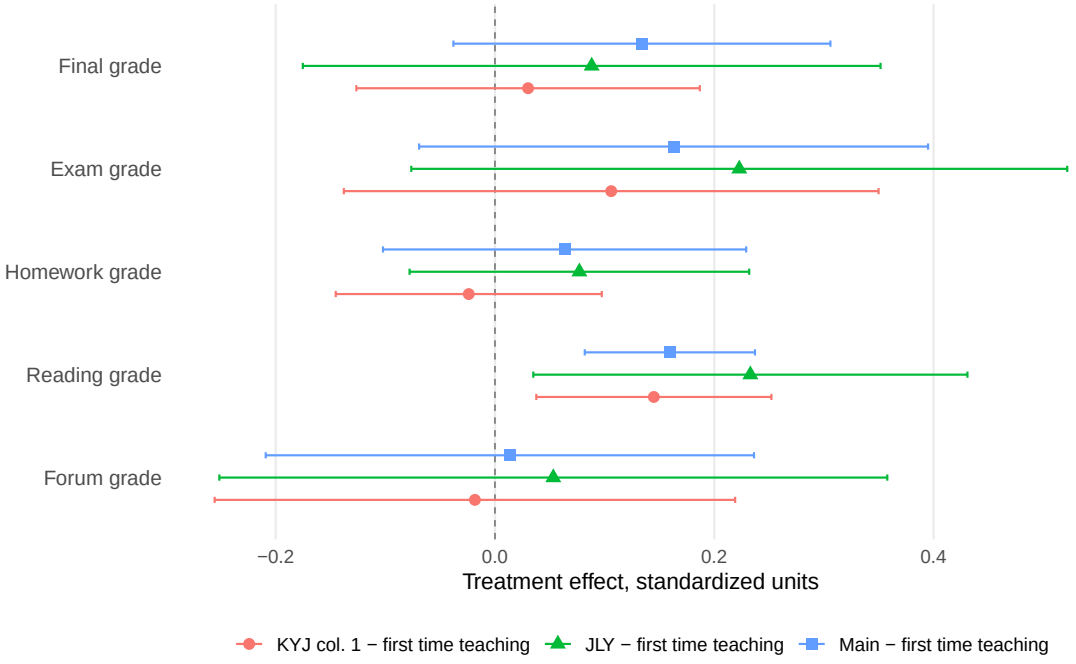

Figure A4. Running main grade specifications on different samples of judges

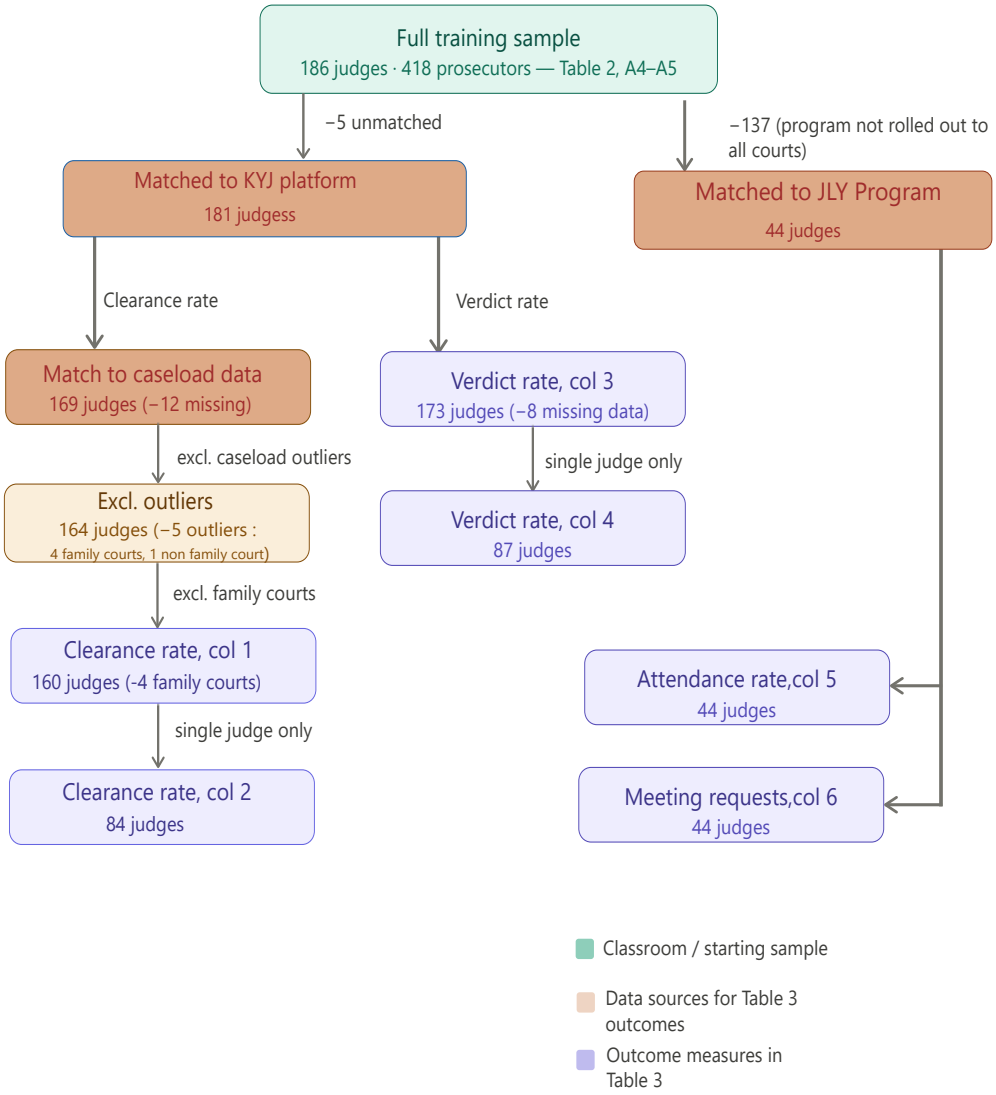

Figure A5. Sample Construction and Data Linkage across Table 3 Outcomes

Table A13. Summary statistics and balance table at the class level (only first-time teaching)

	Summary Statistics						Balance Tests	
	Treatment			Control				
Statistics	Mean	Sd	N	Mean	Sd	N	β	p-value
Class-level stats								
Number of students	26.194	4.072	72	25.871	3.961	62	0.018	0.298
Share of female teachers	0.222	0.419	72	0.210	0.410	62	-0.032	0.759
Share of judges	0.313	0.172	72	0.307	0.178	62	0.019	0.692
Share of prosecutors	0.687	0.172	72	0.693	0.178	62	-0.019	0.692
Share of female students	0.332	0.106	72	0.405	0.090	62	-0.070	0.039
Age	46.370	3.197	72	45.620	3.132	62	0.003	0.243
Years of tenure	5.301	0.737	72	5.440	0.696	62	-0.007	0.399
Years in the bar association	17.843	2.914	72	17.729	2.577	62	-0.001	0.807
Share in criminal court	0.174	0.110	72	0.166	0.114	62	0.013	0.721
Academy's specialist female	0.697	0.463	72					

Note: This table presents balance tests on the monitoring treatment. Sample is restricted to classes where the instructor was teaching for the first time in the program. We present summary statistics displaying means and standard deviations for treatment classes ("Treatment") and control classes ("Control"). Balance tests present an OLS regression of treatment on each characteristic, controlling for strata (participant level and location) and round fixed effects. The regression for number of students is estimated at the class level (N=22) with HC1 robust standard errors. All other variables are estimated at the individual level with standard errors clustered by class. The monitor's gender (Academy's specialist female) appears in the summary statistics columns only and is excluded from balance regressions, as only treated classes have a monitor assigned.

B Appendix: Course Topics

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
d	**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

C Appendix: Course Schedule

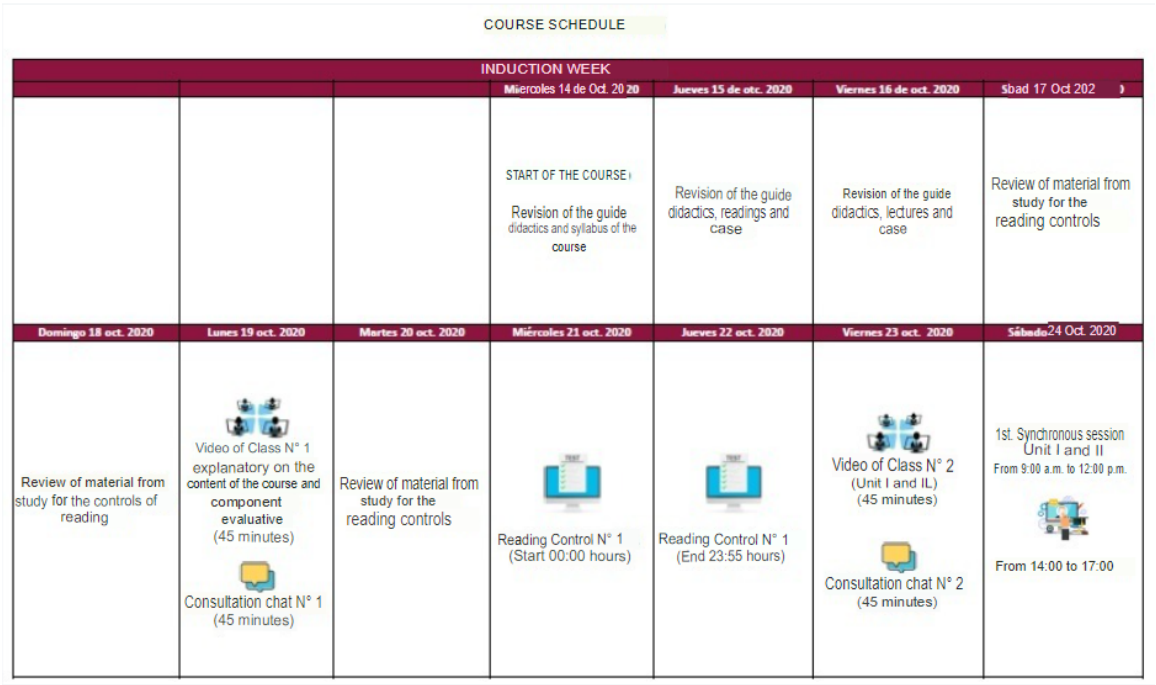

Figure A6. Program structure 1

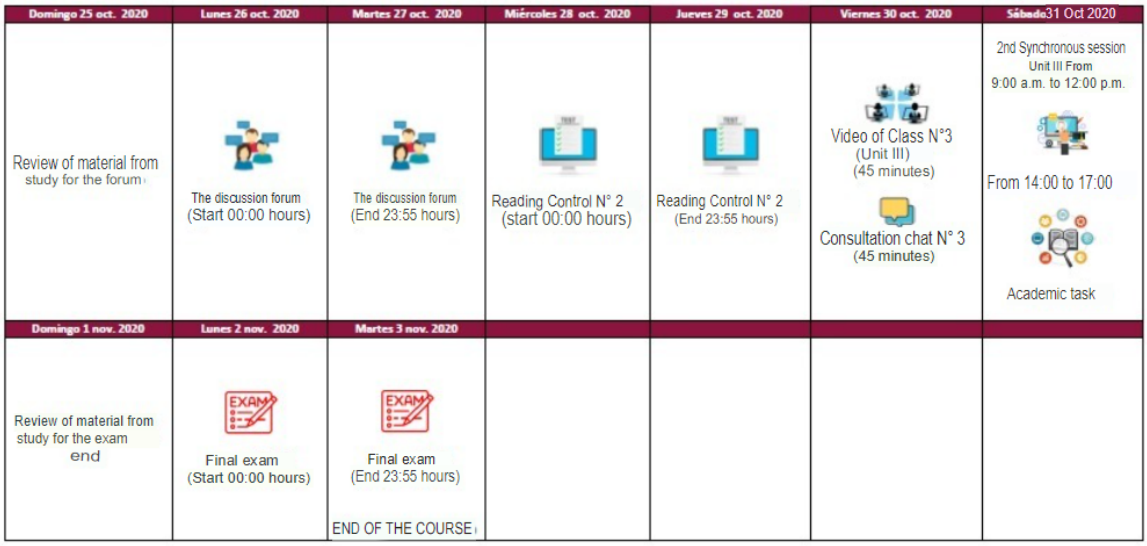

Figure A7. Program structure 2

D Appendix: Observation form

The observation form includes three main sections: 1) Teaching skills; 2) Content mastery; and 3) Class structure. A summary of the characteristics in each section is the following:

1. Teaching skills: Assesses the teacher's enthusiasm, relationship with students, voice modulation, and clarity of language.

2. Content mastery: Evaluates the teacher's ability to clearly and comprehensibly present topics, use real-life examples and analogies, and emphasize key aspects.

3. Class structure:

   (a) "Opening Activities": assesses motivational strategies, communication of session objectives, and student engagement through questions, comments and/or activities.

   (b) "Intermediary Activities": examines the orderliness and relevance of exposition, the integration of case-based examples and related activities, the appropriate use of the syllabus and class materials, the orderly structure of class materials, and student evaluation and feedback.

   (c) "Closing Activities": focus on summarizing session content, evaluating learning outcomes, and providing feedback to students.

CLASS OBSERVATION SHEET

CLASSROOM 11
DATE 16.5.2020

1. **ACADEMIC PROGRAM** : 22 PCA
2. **Course or workshop** : LEGAL ARGUMENTS
3. **Teacher** : XXXX XXXX
4. **Learning Unit** :
5. **Date of synchronous/asynchronous sessions of the course or workshop** : 16/05/2020
6. **Observed session date** : 16/05/2020
7. **Start and end time when the observation is performed** : 9:00 - 13:00 AND 14:00-18:00
8. **Time (start, development, or closure) at which it was observed (indicate what was observed)** : It is indicated according to the lesson plan found in the Guide, indicate what will be observed, according to the Plan.
9. **Monitor or companion** : XXXX XXXX

INSTRUCTIONS

1. The duration of the observation will be 60 minutes (01 chronological hour).
2. Have the lesson plan, which is in the Didactic Guide.
3. The monitor (companion) must observe the actions of the teacher and assign an assessment according to the following scale:

SCALE	DESCRIPTION
A = 20	Outstanding
B = 18	Achieve the expected objectives with satisfaction
C = 15	It is in process.
D = 12	Requires support to achieve goals
E = 8	Presents serious difficulties or can do it and does not

4. If at the time of supervision any of the items cannot be observed, leave blank.

TEACHING SKILLS (HD)		A (20)	B (18)	C (15)	D (12)	And (08)
1.	It transmits interest and enthusiasm towards the class topic.		x			
2.	Relates to students in a cordial and respectful manner.	x				
3.	Its volume, timbre, pitch and voice modulation are adequate.		x			

3.	Carry out the didactic actions according to the times programmed in the Class Plan (See didactic guide)	x				
Process of reflection on what has been worked on						
4.	It promotes the reflection of the students in the class on what has been worked on relating it to their jurisdictional work through questions or activities.	x				
Academic guidance (attention to queries)						
5.	Guide consultations in a clear and timely manner.	x				
Teaching strategies used.						
6.	Uses the case method or other teaching strategy (brainstorming, role play, directed debate, among others) for the development of the content or learning activities of the session.	x				
7.	Develop cases or other activities that allow to strengthen the subject worked.	x				
8.	Follow the procedure for the use of the teaching strategy (case method, brainstorming, role play, directed debate, among others)		x			
Study materials (syllable, teaching guide, or others) and resources						
9.	Use the syllable and the learning activity tutorial for the development of your class.	x				
10.	Use resources (videos, whiteboard, or other) to facilitate learning	x				
Organization of work in the classroom						
11.	Perform individual and/or group activities in the learning session.	x				
12.	Give clear indications to carry out the work whether individual or group.	x				
13.	Provides adequate and timely feedback to students, when performing some group or individual work, pointing out the difficulties encountered and positively valuing the successes.		x			
Learning assessment and feedback						
14.	Evaluates students through the use of cases or other activity that allows to see the application of what they learned in class to their jurisdictional work.	x				
15.	Use rubric or homework analysis sheet or other assessment instrument to grade the work of the students.					
16.	Provides adequate and timely feedback to the students, pointing out the difficulties encountered and positively valuing the successes.		x			
Average HC = (Total AD/16)=19.6						

CLOSURE ACTIVITIES (AC)		A (20)	B (18)	C (15)	D (12)	And (08)
1.	Perform closing activities: ask the class group to indicate ideas strength of what was developed in the session, and / or evaluate the results through a final exam or the elaboration of an academic task, etc.		x			

4.	The language used is clear and intelligible.	x				
Average HC = (Total HD/4) = 19						
CONTENT MASTERY (DC)		**A (20)**	**B (18)**	**C (15)**	**D (12)**	**And (08)**
5.	Develop topics in a clear, orderly and understandable manner.	x				
6.	It uses examples from the work context and relates them to the contents developed.	x				
7.	Explain the topics using examples, analogies, etc. that facilitate understanding.	x				
8.	It emphasizes the most important aspects of the subject he works.	x				
Average HC = (Total DC/4)=20						

MOMENTS OF THE SESSION -FORMATIVE PROCESS (M)						
INITIATION ACTIVITIES (AI)		**A (20)**	**B (18)**	**C (15)**	**D (12)**	**And (08)**
Motivation and initiation						
1.	Perform a motivational activity at the beginning of class (anecdote, example, dynamics, video, etc.) related to the topic of the learning unit you will be working on.					
2.	Communicates the capacity to be achieved in the session relating it to the usefulness for its jurisdictional work.					
3.	It collects and uses the previous knowledge (knowledge they may have regarding what is addressed) of the students to build the new learning, through previously selected activities and strategies.					
4.	Encourage the student to participate constantly through clear questions, comments, or activities.	x				
Average HC = (Total AI/4)=20						

MOMENTS OF THE SESSION -FORMATIVE PROCESS (M)						
DEVELOPMENT ACTIVITIES (DA)		**A (20)**	**B (18)**	**C (15)**	**D (12)**	**And (08)**
Development of the learning session						
1.	The presentation is orderly, precise and relevant.	x				
2.	The selected examples and/or cases are addressed according to the level (I, II, III or IV) that corresponds to them, if applicable, and the place (Lima or province) where they work.	x				

2.	It uses assessment instruments to rate the student's learning.					
3.	Provides adequate and timely feedback to the students, pointing out the difficulties encountered and positively valuing the successes.	x				
Average HC = (Total AC/3)=19						

CONCLUDING REMARKS

Very well the performance of the teacher in the development of the session, the initiation activities should be considered to optimize their performance.

E Appendix: Instructor feedback

The design of the intervention goes beyond simply observing instructors and ensuring they do their job. Personalized feedback, in addition to monitoring, facilitates the sharing of experiences and the targeted improvement in the delivery of the class. Below is an example of the type of feedback instructors receive and the actions instructors agree to take to improve their teaching.

Instructor Feedback Example

Strengths:

- Highlighting the teaching skills transmitting interest and enthusiasm about the subject, the respectful and cordial relationship with the students, the volume, timbre, tone and modulation of the voice, and the use of clear and intelligible language.
- Mastery of the subject is highlighted: the use of examples from the labor sector to facilitate understanding and the linkages made with the content taught.
- The development of the initial activities: the presentation of the activities to be carried out.
- The presentation is clear, precise and relevant.
- Provides adequate and timely feedback to students.

Agreements / commitments to improve:

- The use of the camera when there is student participation to ensure human interaction and meaningful learning.
- Clarification in the reasoning. State the skill that will be addressed in the session.
- Carry out closing activities: ask the class for the main ideas from the session and the instructor complements the class contributions.
- Use some of the tools on the Blackboard platform.

ACT OF AGREEMENTS AND COMMITMENT

1. ACADEMIC PROGRAM		: TRAINING PROGRAM FOR PROMOTION
2.	**course or workshop**	: Plenary Agreements of the Supreme Court of Justice in Matter Civil
3.	**teacher**	[redacted]
4.	**Assigned classroom**	: 30
5.	**Observed learning unit:** Learning Unit No. 01 and 02	
6.	**Observed session date**	: Saturday, October 24, 2020
7.	**Date and time of beginning and end in which the feedback and the Minutes**	: October 29, 2020 8:00 p.m. - 8:30 p.m.
8.	**Monitor or companion (Specialist in the area of Methodological Design)**	[redacted]

1. SELF-ASSESSMENT OF THE TEACHER

Strengths:

- Motivation in the students for the participation and attention to the synchronous session.
- The teaching experience of the teacher in synchronous work means that the sessions can be be more dynamic.

Aspects to Improve:

- Improve time management in student participation.
- Promote the participation of all students.

2. SCOPE OF WHAT WAS OBSERVED

Strengths:

• The teacher has a lot of management of information sources. • The teacher's experience in the development of the topics covered means that the students can answer all their questions with the course teacher.

Aspects to Improve:

- Increase motivation in the subject matter, to achieve more participation of the entire cluster.
- In order to improve access to information sources, the search for readings will be promoted. more updated.

ACADEMIA DE LA MAGISTRATURA
ACADEMIC DIRECTION

3. AGREEMENTS AND COMMITMENTS

AGREEMENTS	COMMITMENTS
• The teacher agrees that the most active participation of all magistrates registered in the classroom should be promoted. • The teacher will promote the use of more sources of information.	• In the next session, we will seek to motivate the students more in order to achieve the participation of the majority of the students. • It will seek to take more care of the participation times of the students, so that everyone can participate.

OBSERVATIONS:

• The teacher proposes that: During the development of the course, all the jurisdictional plenary sessions could be analyzed and, in particular, analysis of the last plenary session. Because of its importance.

• Regarding the X Full Casatorium on the Examination, the teacher proposes that a special academic course or activity should be carried out in order to further promote their knowledge among magistrates and magistrates

[redacted]
(Methodological monitoring team)

[redacted]
activity teacher

F Appendix: Email to announce the treatment to the instructor

Email sent to each instructor in the Promotion Program of the Judicial Academy who taught classes in the treatment group:

Dear Dr. [Name],

Instructor of the Legal Interpretation and Argumentation Course,

I would like to take this opportunity to extend a cordial greeting, thank you for your valuable participation as an instructor in the Promotion Training Program, and inform you that, within the framework of the Regulation of the Professional Regime that provides teaching services to our institution and the Pedagogical Innovations being implemented by the Academy of the Judiciary, an impact evaluation of this training process is being carried out this year, with the support of the World Bank and our methodologists. This activity includes the monitoring and methodological support of the teaching work in the courses that are part of the 22nd APP, with the aim of providing guidelines and methodological orientations to help instructors achieve optimal performance in synchronous sessions, ensuring academic excellence and service quality.

In this regard, the methodological monitoring team of the Academic Directorate of AMAG will enter your videoconference room in the two synchronous sessions, and we request your valuable support in this matter. After this, they will contact you to provide feedback on your performance.

I would appreciate it if you could facilitate the development of both activities (monitoring and feedback), all with the aim of optimizing your performance as an AMAG instructor.

We wish you much success.

Sincerely,

Hipólito M. Rodríguez Casavilca

Academic Director, Judicial Academy of Peru