Learning to constrain: Political competition and randomized controlled trials in development*

Cristina Corduneanu-Huci† Michael T. Dorsch‡ Paul Maarek§

December 10, 2017

Abstract

This paper provides a political economic analysis of impact evaluation experiments conducted in international development. We argue that in more politically competitive environments, where incumbents face a higher probability of losing power, governments have stronger incentive to run Randomized Controlled Trial (RCT) experiments to constrain successors' margin of policy discretion. Moreover, the effect of competition on the probability to host RCTs is stronger in more polarized societies since the incumbent's cost of losing power is higher. We first propose a formal model and then empirically examine its theoretical predictions using a unique data set on RCTs that we have compiled. Over a panel of Indian states and a cross-national panel, we find that certain RCTs are more likely to occur in electorally competitive jurisdictions, and that the effect is amplified by political polarization. We demonstrate that politics matter for when, where, and with which partners RCTs in development happen.

Keywords: Program evaluation, RCT, External validity, Political accountability, Political competition, Development policy

---

*We thank Dragana Marinković and Armine Hakhinyan for their excellent research assistance. Participants at the Silvaplana Workshop on Political Economy provided exceptionally useful comments on an earlier draft of this paper.

†Central European University, Nádor u. 9, 1051 Budapest, Hungary; corduneanu-hucic@ceu.edu.

‡*Corresponding author.* Central European University, Nádor u. 9, 1051 Budapest, Hungary; dorschm@ceu.edu.

§Université de Cergy-Pontoise, 33 boulevard du Port, 95011 Cergy-Pontoise Cedex, France; paul.maarek@u-cergy.fr.

1 Introduction

Randomized control trials (RCT) as the most rigorous form of impact evaluation have been increasingly used to assess social policy programs and have become critical for scaling up projects in international development (Duflo, 2004). Recently, field experimentation has also expanded in the areas of political institutions and rendered important insights into individual and organizational incentives that may lead to better governance outcomes (Humphreys and Weinstein, 2009). The clear promise of RCTs in terms of policy learning could be enhanced by solutions to the problems of external validity, cumulative knowledge, and general equilibrium theory generation (Green and Gerber, 2002; Humphreys and Weinstein, 2009; Acemoglu, 2010; Banerjee and Duflo, 2014). In this spirit, some argue that the accumulation of enough experimental findings and the availability of internally valid evidence across settings or obtained through selective trials might be able to lead to generalization (Chassang et al., 2012; Dehejia et al., 2015), while others propose to study pre-specification and coordinated research involving field experiments as a way out of the external validity trap (Dunning, 2016; Dunning et al., 2017).[1]

This paper contributes to these efforts by taking one step back and examining the political economic incentives that shape the demand for RCTs. How may governments strategically use RCTs beyond their intended purpose of policy learning? In what contexts are impact evaluations more likely to occur and why? Is there variation across policy areas? We argue that politically competitive environments are more favorable toward risky experimentation given that politicians have incentives to fend off criticism, secure bureaucratic buy-in, and embed policy constraints on opponents if and when they lose elections. Our theoretical focus is on the incentive for incumbent governments to reduce the extent of discretion in policy-making, an incentive that we argue is stronger in more competitive political environments. Moreover, we provide a formal model and an empirical analysis to demonstrate how domestic politics matter for when and where RCTs concerning development policies occur. By analyzing how political incentives may shape decisions to run an experiment, our paper introduces the possibility of a "political site selection bias" into the debate about generalizability of RCT findings for international development policy.

Experiments entail both political benefits and costs for politicians and bureaucrats. Beyond the direct bureaucratic cost the experiment involves, if the tested government program is a key component of the incumbent's policy platform, findings pointing at its effectiveness

[1] For political economy and governance experiments in particular, several proposals attempt to combine theory and RCTs in order to shift from partial to general equilibria explanations, deal with systemic complexity of policy interventions, and uncover causal mechanisms (Acemoglu, 2010; Deaton and Cartwright, 2016; Dunning, 2016; Wantchekon and Guardado R, 2011).

help garner public support and credibly respond to critics. In contrast, null or negative results might prevent politicians from implementing their preferred policies and cater to core constituencies. Even evaluations conducted by NGOs or research institutions oftentimes require formal or informal interactions with the government, expose a large share of citizens to an intervention, and may offer and publicize a superior alternative compared to the policies implemented by the incumbent. We argue that these processes have their own political stakes that are likely to affect the occurrence and implementation of an RCT, with implications for how well experimental results travel across various contexts and time.

More specifically, in this paper we argue that the incentive for an incumbent government to run an experiment may be shaped by the degree of electoral competition and the resulting probability for the incumbent to lose power. We develop a simple theoretical model that focuses on experiments that concern policies that introduce more constraints on the discretion of governmental decision-making. Reforms that "tie the hands" of the government impose utility losses on future incumbents as constraints reduce policy discretion to capture rents. On the other hand, the reforms benefit future opponents since constraints limit the ability of the incumbent to colonize bureaucracies following elections, make targeted transfers to his own group, and oppress opposition groups using discretionary executive power. In this context, the incentive to conduct such a reform increases when the incumbent faces a high probability of losing power since he will have a higher probability to benefit from the constraints (as an opponent) and a lower probability to bear the costs of those constraints (as an incumbent).

In our model, there are two motives for running an experiment before undertaking a reform. First, the experiment allows for a better design of the reform, which increases the probability that the policy change produces a sizable effect on constraining discretionary decision-making. Second, the magnitude of the effect of the constraints on the government may be uncertain and the experiment allows the incumbent to know the effect of the constraints with more precision, which is valuable information when deciding wether or not to implement the policy change after the experiment occurs. For both motives, the benefits to the incumbent of running an experiment are increasing in the degree of political competition.

We characterize different equilibrium configurations. In some cases, the incumbent runs an experiment prior to a reform only for a sufficiently high level of political competition. In some other cases, the incumbent never pursues an experiment. The main result of the model is that the probability to observe an experiment is greater when the probability of losing power is higher. Furthermore, we show that the effect of political competition is magnified by political polarization since the difference between the policy preferences of the incumbent and opposition are greater in more polarized societies.

In order to test such predictions, we make use of an original hand-coded dataset of 289 RCT experiments conducted in 43 countries between 1995 and 2015 that draws on RCTs included in the online repository of the Abdul Latif Jameel Poverty Action Lab (J-PAL). We have categorized the experiments according to the type of development policy that is being tested, ranging from bureaucratic accountability to minority inclusion. We have merged this data with a variety of political and economic indicators that we use as explanatory variables to analyze when and where policy evaluations happen in a cross-country panel regression framework. As many experiments have occurred in India, we have also compiled an analogous panel dataset that covers Indian states. We present results both for the cross-country panel and the Indian state panel.

Using proxy variables for the probability of experiencing a political turnover, we first show that countries where political incumbents face a higher probability of losing power have a higher probability of running experiments. The effect is most robust for experiments that focus on public sector accountability and absent for some other categories of experiments, such as electoral learning and minority inclusion. Second we show that the effect is only observed when the government is involved as a partner in the experiment, which the political economic analysis at the base of our paper. Non-governmental partners such as NGOs, universities or donors are not expected to react to such political incentives and, indeed, on average they do not. Finally, we show the effect of political competition is much higher in ethnically fragmented societies, which we use as a proxy for polarization in the developing world. Taken together, the results support the "tying-the-hands" mechanism at the center of our theoretical analysis.

The results clearly indicate that some categories of experiments we observe occur in particular contexts, implying that the site in which the experiment occurs and actors involved in the experiment are not randomly chosen. More generally, this paper introduces a first analysis to understand the incentives for a politician/bureaucrat to run an experiment which is a crucial step to start understanding why experiments occur in some places rather than others. The next section presents a literature review focusing on the politics of experimentation in international development. The third section presents a formal political economic model and derives testable hypotheses. The data is introduced in the fourth section and the fifth section presents our empirical analysis. The final section concludes and offers some avenues for future research.

2 Literature review

Recent contributions to the literature on RCTs in international development increasingly acknowledge and debate the complex political and bureaucratic incentives to conduct policy experimentation, as well as their implications for external validity (Banerjee et al., 2017; Deaton, 2010; Imbens, 2010). Ex-ante attitudes of policymakers toward evaluation vary widely across contexts and reflect risk-taking. For large scale interventions that cover a significant share of the voting population of a district, the very nature of the treatment, be it in cash, kind, or altered institutional rules, becomes a political stake for the officials directly or indirectly involved in its implementation. Therefore, deciding to sponsor, run, or host a policy experiment entails both costs and benefits whose analysis is crucial for understanding the specific contexts that are more likely to foster policy learning. We discuss such costs and benefits in turn.

2.1 Political economic costs of hosting RCT experiments

First, evaluation triggers significant ideological and credibility costs for politicians and bureaucrats because of the inherent uncertainty of results. If the findings of a randomized trial match the previously advertised preferences, then they benefit the policy-maker. The first quasi-RCT ever implemented in social policy, evaluating the introduction of a negative income tax in the state of New Jersey in 1968, was "used and misused" by the Nixon administration to promote a legislative reform (Gueron and Rolston, 2013). If the results, however, depart significantly from ideal points, they may undermine ideological cues, party positions or particularistic rents and transfers to constituents, by tying the hands of politicians who want to implement vote-winning agendas instead. One related concern, often expressed by bureaucrats, is that politicians do not want to convey bad news to their voters in case their championed program failed to demonstrate impact. During the early 1990s, for instance, most United States governors opposed randomized evaluations of heavily contested welfare programs because of the fear that the results would depart from campaign promises. According to one evaluator involved in these debates, the introduction of mandatory randomized policy trials "... at least meant that governors who were trumpeting success would ultimately have to face the reality of hard evaluation findings" (Gueron and Rolston, 2013). Even evaluations that find a positive impact of a policy, yet more modest than initially signaled to voters, entail political credibility costs and antagonize various stakeholders such as evaluators, program managers and politicians with conflicting interests (Lipsky et al., 2007). During the initial RCTs pioneered in the United States in the 1970s to evaluate ideologically contested welfare-to work programs, while researchers perceived randomization as the only methodologically sound way to distinguish the impact of a program from background factors, politicians feared that reporting more modest claims of impact would put them in an inferior position compared to their peers that took political credit for broader outcomes of programs such as the overall employment rate of welfare beneficiaries (Gueron, 2017). In the case of deeply polarized areas such as the military intervention in Afghanistan, some United States legislators argued that the demand for instant impact results became an impediment for the medium and long term goals they had in mind when supporting the operation and communicating it to their constituents (Lawson, 2012). For bureaucrats, impact evaluation is also inherently political since its potentially negative or null effects may determine funding and the survival of a program, and translate into significant career costs (Pritchett, 2002; Weiss, 1993). In the words of one development practitioner, "if you don't ask [about results], you don't fail, and your budget isn't cut" (Lawson, 2012).

Second, evaluations that aim at increasing the accountability of decision-makers trigger the highest costs and often face resistance from bureaucrats or politicians who feel directly scrutinized and anticipate losing rents. In fact, in many countries, the demand for evaluations correlates inversely with political rents across issues. For instance, in African developmental patrimonial states such as Rwanda or Malawi, RCTs are conducted in health and agriculture where elites are genuinely interested in learning policy effects, but not in more politically sensitive areas (Porter and Feinstein, 2014). In a well-known Indian example, a large scale RCT conducted by JPAL in partnership with the Government of Bihar aimed to assess whether the introduction of electronic fund-flow management and the elimination of payment intermediaries would reduce the overall level of corruption and leakage detected in the implementation of India's largest workfare program, Mahatma Gandhi National Rural Employment Guarantee Scheme (MGNREGS) (Banerjee et al., 2016). During the period of implementation of the RCT in the treatment group (2012-2013), program officers at the district level became increasingly displeased with the experimental changes in funds disbursement on grounds of increased administrative burden, unfamiliar IT technology, and a reduction in past rents such as ghost workers and other forms of leakage. Since the field experiment was perceived as costly from an administrative and political standpoint, most district officials did not support it and lobbied higher decision levels to stop it. As a result, the Government of Bihar eventually decided to discontinue the intervention in spring 2013 (Banerjee et al., 2016).

Third, the randomization processes involved in impact evaluations may impose electoral costs on politicians because they entail withholding treatments or benefits from a large segment of voters that happen to be assigned to the control group. For instance, in 1989, the legislature of the state of Florida almost banned control groups because of political opposition to the randomization used in testing Project Independent, a welfare-to-work program (Gueron, 2017). Similarly, an independent team evaluating the largest Colombian conditional cash transfer program, *Familias En Accion*, encountered significant design challenges because of the opposition of the Colombian President facing elections in 2002. The randomization of beneficiaries according to the strategy envisaged by evaluators would have involved a slower expansion of the program in municipalities designated as a control group. Because of the anticipated unpopularity of the experimental design in an election year, the President vetoed the strategy of randomization (Briceño et al., 2011).

In the formal model we develop, such direct and indirect costs of policy experimentation are considered in a reduced form only. Despite costs, hosting an evaluation may bring significant benefits to a politician or a bureaucrat. We discuss three paths through which hosting field experiments may translate into immediate or future political gains.

2.2 Political economic benefits of hosting RCT experiments

First, in line with the ignorance thesis, evaluations help decision makers gain knowledge regarding what policies are more effective. In well governed contexts, some politicians and high level bureaucrats willing to signal technocratic competence and foster learning have either designated special units of evaluation embedded within the administration, or have signed comprehensive memoranda of understanding with internationally reputed researchers governing their commitment to evidence-based policy making. For instance, in 2014, the United States White House under President Obama created a Social and Behavioral Science Team with the mandate to conduct rigorous impact evaluations of government programs. Similarly, the Indian state of Tamil Nadu entered an agreement with JPAL in 2014 because of a Chief Secretary committed to evidence-based development programs. Whereas this type of learning brings significant benefits to decision-makers, it may also trigger policy specific costs contingent upon the distance between ideological stances or party positions and the experimental results.

For a subset of development RCTs, learning comes with political benefits and no indirect costs for politicians who want to refine the signals they send to voters, test the most effective ways to communicate with their platforms to the electorate, and increase their vote share. In the United States and Canada, it is common for campaign offices of political parties and candidates to partner with researchers in order to understand what mechanisms of contact and communication work best for voter mobilization (Green and Gerber, 2015; Loewen et al., 2010). Developing contexts with a lived practice of policy innovation foster similar learning incentives. In a classic political economy RCT conducted during a presidential campaign in

Benin in 2001, parties were genuinely interested in learning whether programmatic campaign messages work with their constituents (Wantchekon, 2003). In the empirical section, we label this type of randomized experiments as Electoral Learning and treat it as a distinct category of evaluation that, in stark contrast with accountability learning, bears virtually no career costs for politicians or bureaucrats. Because of this unique feature, we anticipate that electoral learning RCTs are less sensitive to context.

Second, if favorable, the results of RCTs conducted by reputable independent organizations convey the bureaucrats and politicians' intentions to implement policies based on rigorous evidence, address criticism, and establish objectivity in politically contentious and rhetorically spun issue areas (Gueron, 2017). This strategy is particularly effective in in the case of polarized issues and environments where the support for a program is short-lived in the absence of evidence that the opposite factions may recognize as neutral. In the words of one of the pioneers of RCTs assessing politically contested United States welfare programs since the 1970s, "the only way to get out of the pickle of these dueling unprovable things ... and salvage the credibility of the program ... was to get an evaluation of unquestioned quality and go forward" (Gueron and Rolston, 2013). In such cases, respected impact evaluation studies help design a better policy and signal to the opposition its effectiveness if adopted.

For instance, in India, the government promoted the introduction of *Aadhaar*, or biometric smart cards, as one of the most ambitious governance programs designed to reduce leakages in social benefit transfers. Its implementation has triggered significant backlash. For program designers, the biometric card meant increasing efficiency in service delivery, reducing corruption and improving bureaucratic accountability. For political opponents and civil society critics, the introduction of this tool in parallel with the elimination of the previous mode of delivery raised question related to government surveillance, potential discrimination of the poor and de-facto exclusion of vulnerable populations such as the elderly and the disabled from receiving any benefits at all. In this polarized context, some politicians championed an RTC conducted by JPAL researchers that demonstrated the high impact of biometric cards on leakage reduction in order to provide evidence and advance a program in line with their platform for upcoming elections. (Muralidharan et al., 2016).[2]

Another example is provided by Liberia which made international headlines because of the plan to unprecedentedly outsource its entire pre-primary and primary education system to an international for-profit company. Introduced by the Minister of Education in 2015 and

[2] See also: "Political pressure halted direct benefits transfer for LPG" Karthik Muralidharan Business Standard. 2016 Available at: http://www.business-standard.com/article/economy-policy/political-pressure-halted-direct-benefits-transfer-for-lpg-karthik-muralidharan-114040400144_1.html.

championed politically by President Sirleaf Johnson, Partnership Schools for Liberia (PSL) aimed at addressing the crisis of an education system plagued by chronic teachers absenteeism, and characterized by one of the world's worst school attendance and literacy rates. The program deeply polarized the minister, the legislature, the main teachers' union, the UN Rapporteur on the Rights to Education and other stakeholders, around both ideological issues regarding private versus public education, as well as around concrete consequences affecting teachers' union bargaining power, unsustainable costs and, for some, questionable educational quality.

In this context, in 2016, the Minister and his advisers commissioned an RCT study with rigorous randomization conducted by an internationally recognized team of evaluators from Center for Global Development and IPA. The RCT results allowed the Ministry to appease heated public criticism, as well as to lock in the program before the 2017 election where the incumbent was very likely to lose power. One year later, recalibrating the political costs of gathering additional evidence, the same Minister blocked last minute an US-based evaluation team that planned to collect qualitative data to supplement the quantitative findings of the RCT, and attempted to commit to a larger scale-up for the second year, diverging from the recommendations of the RCT study, and triggering a corrective public reaction of the researchers involved (Adamson 2017, Sieh 2016).[3]

In both the Indian and the Liberian examples, the RCT evaluation studies helped politicians provide evidence to the opponents and voters that the policies they proposed generate a positive impact, in an attempt to make controversial reforms more accepted in contested environments. In these cases, political polarization gave a strong incentives to policy makers to commission experiments.

Third and relatedly, decision-makers who anticipate political turnover following the next elections have strong incentives to use evaluation studies to build credibility, signal effective design, and insure policy legacy. Indeed, in polarized political environments, it is very likely that most programs will not survive political turnover and policies will be sharply reversed. If an RCT generates supportive results, it may secure buy-in from bureaucrats, voters and some opponents, increasing the chance that the policy is continued under the next administration. If we think of evidence-based policy making as a key component of competent and apolitical bureaucracies, this argument is in line with previous theories emphasizing how electoral uncertainty, or the likelihood of losing elections, motivates parties and politicians to insulate

[3]Need BibTeX citations: Adamson, Frank. 2017, June 15. "Liberia: Open Letter to George Werner, Minister of Education Liberia." Available at: http://allafrica.com/stories/201706150861.html. and Sieh, Rodney. ' 'Is Bridge Bullying Liberia Into Submission? Liberia's Education Outsource Plan Dilemma." Front Page Africa. Available at: http://frontpageafricaonline.com/index.php/politics/1409-is-bridge-bullying-liberia-into-submission-liberia-s-education-outsource-plan-dilemma.

bureaucracies in order to prevent being punished by competitors in the future through these organizations (Moe, 1989; Grzymala-Busse, 2007; Geddes, 1994; De Figueiredo, 2002).

RCTs have provided effective advocacy tools for insuring the survival of social policies and the legacy of politicians in very uncertain contexts. The most celebrated conditional cash transfer program that eventually became the source of inspiration for the diffusion of CCTs around the world, *Progresa*, was adopted in Mexico in 1997 during an uncertain time of political transition and democratization. President Ernesto Zedillo (1994-2000), its initiator, attempted to distance his administration from the corruption scandals that plagued previous cabinets of the hegemonic party rule. Restoring confidence in the bureaucracy became a cornerstone of his office in the background of a transition to multi-party competition. In this political context, Progresa's design benefited from the early integration of impact evaluation conducted by a US based research institute, and was essential for signaling to voters the apolitical and technocratic nature of social programs, a key component of Zedillo's policy platform under the "New Federalism" (Faulkner, 2014). Despite its imperfections, the initial success of the RCT revealed a strong impact of the CCT program. The final reports on the experiment were released and widely publicized in 2000 during presidential elections in order to signal to opponents and voters a break with the heavily politicized anti-poverty programs of PRI's past and increased accountability. Despite Zedillo's electoral loss, the publicity surrounding the RCT insured that the new administration will continue the program (Levy, 2006). In fact, partly because of the impact demonstrated by the study, Progresa was the first social program to survive a presidential cycle and the first alternation in power in Mexican politics in seventy-one years (Faulkner, 2014). Unlike the Colombian *Familias En Accion*, impact evaluation served as an advocacy tools to signal the de-politicization of social policies, increase accountability, and insure legacy.[4]

Similarly, Indonesia's national rice distribution pilot (Raskin) and US federal social programs such as Head Start, or the Women, Infant, and Children nutrition program or welfare programs, built credibility and garnered the political support of the public despite polarization because of strong evaluation results (Gueron, 2017; Lipsky et al., 2007). In all these cases, the policy environment was crucial for understanding the decision of policy-makers to commission evaluation studies. The incumbent's high probability of losing power rendered programs vulnerable. Therefore, beyond policy learning, the experiments aimed at securing continuity.

[4] Note that the two types of benefits were different in nature. In the first case, the benefits of the experiment occurred at the policy design stage, before implementing a large scale policy change. In the second case, such benefits occurred ex-post, after policy implementation, as the study aimed at convincing political successors that the policy was well designed. In our theoretical treatment we will focus primarily on the first motive.

In conclusion, anecdotal evidence implies that the political environment impacts significantly the decision makers' costs and benefits of commissioning or running an RCT evaluation study. Accordingly, we expect that the probability to observe it will be endogenous to both context and the benefit cost ratio associated with specific categories of policy experiments, and articulate the theoretical channels of causation. Understanding why and in what cases site and partner selection bias occurs is crucial for disentangling the treatment effect from the general context where the evaluation takes place (Allcott, 2015; Allcott and Mullainathan, 2012; Blair et al., 2013; Bold et al., 2013; Grose, 2014; Vivalt, 2016). We hypothesize that certain types of evaluations are more prone to selection bias than others and focus our theoretical and empirical investigation on classifying political /bureaucratic costs and benefits of learning and the conditions that foster them. The paper argues that in particular, RCT experiments evaluating accountability (either of elected politicians or public sector employees) are prone to site and partner selection effects because their probability of occurrence depends on political competition and polarization.

3 A formal theory

3.1 Overview of the model

In our model, the incumbent faces a risk of loosing power and having the opposition in office for the next period. Without any constraint on the discretionary use of power, the opposition experiences some utility loss. On the other hand, when there exists some constraint on power, it is the incumbent government that bears a utility loss due to the fact that the incumbent is not able to exercise power with discretion. The trade off is clear: if the incumbent introduces some constraints on power in the next period, he experiences a utility loss if re-elected, but a utility gain if replaced. Political competition affects the probability of losing power and thus shapes the incumbent's incentive to implement a reform that constrains future power.

In our model, there are two related motives for running an experiment. First, we assume the policy change may not yield effective constraints on power in the next period, as changing institutions to modify the use of power is complex and the policy change may not have bite. In our simple model, running an experiment allows for a better design of the reform which increases the probability the policy change actually constrains power. Second, the effect of constraints on power may be uncertain from the point of view of the incumbent. We suppose that the experiment improves the incumbent's precision of the expectation of the effect of the reform, which helps to decide whether or not to implement the policy change. As described in the literature review, experiments also entail some costs (direct administrative costs or more indirect political costs). Here again, political competition will affect the balance between the benefits of running an experiment (which are higher when political competition is more intense) and the associated costs.[5]

3.2 Environment

There are two periods denoted by $t = \{1; 2\}$ and two parties (politicians) denoted by $j = \{I; O\}$ for the incumbent and the opponent, respectively, who discount the future at rate β. In this simple model, there are no voters and the incumbent faces an exogenous probability of loosing power ζ (as in Besley et al. 2016) which is common knowledge and which corresponds to the degree of political competition. We assume that the incumbent enjoys a rent R^I while in office, which can correspond to an ego rent or to any benefit a politician may derive from holding office. We propose a very reduced-form manner to express how constraints on power affect the utility of the politician. We assume that if there are constraints on power, the incumbent has a utility loss of C_I each period. In Besley et al. (2016), such a utility loss follows from the ruler's reduced ability to make targeted transfers to his own group. But it could also correspond to some limitations on the ability of the ruler to derive some personal benefits from his position.On the other hand, the opposition derives a utility gain of C_O from constraints on power. In Besley et al. (2016), such a utility gain to the fact that the incumbent is not able to favor its own group through targeted transfers which makes the group the opponent belonging better off. This can also corresponds to the fact the civic rights are better guaranteed when there exists some more constraints on the executive. We assume that this benefit can take two values: a high value C_O^h with probability h and a low value C_O^l with probability $(1 - h)$. The value of C_O is unknown to the incumbent.

The incumbent has two sequential policy choices. First, he can decide to run an experiment before introducing the policy change $E = \{0; 1\}$. Second, he can decide to undertake a policy change $R = \{0; 1\}$ to implement some constraints on power the next period (in $t = 2$). There are four policy combinations $P(E; R)$ as a result.

When reforming without running an experiment the policy change yields effective constraints on power with probability P_R. When an experiment is introduced before the policy change is implemented, the policy change has a probability P_E of effectively constraining power, with $1 > P_E > P_R$. This is due to the fact that running an experiment allows to have a better design of the policy which is more effective in reaching its goal. Even with an

[5] Note that our model and the motive for running an experiment can be interpreted very broadly in terms of "tying the hands" of future governments. It could correspond to limiting the ability of the government to make targeted transfers. But it could also be a tool to "lock in" the incumbent's policy in future periods if the experiment yields positive results.

experiment, there is still some uncertainty that the reform will be effective in constraining power. Additionally, when running an experiment, the incumbent learns the value of C_O. Thus, after having choose to run an experiment, the incumbent can choose whether or not it is profitable to pursue a policy change given the costs and benefits that have been revealed. Here, the benefit to run an experiment is that it gives some extra information about the cost and benefit to undertake a reform.[6] Furthermore, running an experiment has a resource cost of c^e.

The timing of the game for the first period is as follows:

1. Nature decides the value of C_O which is high with probability h and which is unknown for the incumbent.
2. Nature decides the value of ζ which is known for the incumbent.
3. The incumbent decides whether or not to run an experiment, $E = \{0, 1\}$.
4. The incumbent decides whether or not to implement the policy change, $R = \{0, 1\}$.
5. The election takes place and the incumbent is re-elected with probability ζ.

3.3 Analysis

3.3.1 The post-experiment decision

We analyze the game recursively, starting with the decision to implement a policy change after an experiment. First, consider the value function for the incumbent of *not* implementing the policy change following an experiment:

$$V_I\left(R = 0; E = 1\right) = R^I + \beta\left[\zeta R^I\right] - c^e \tag{1}$$

The incumbent is re-elected for a second term with probability ζ and enjoys the rent from being in office the next period. With probability $1-\zeta$ the incumbent will be in the opposition group for the next period and his utility is normalized to zero when there are no constraints on power. Now consider the value function for the incumbent when implementing the policy change following an experiment:

$$V_I\left(R = 1; E = 1; C_O = C_O^i\right) = R^I + \beta\left[\zeta\left(R^I - P_E C_I\right) + \left(1 - \zeta\right) P_E C_O^i\right] - c^e, \tag{2}$$

[6]One could imagine some other information the experiment gives to the incumbent that affects the incentive for pursuing the policy change. It could concern C_I, the utility loss the next incumbent will experience from the constraints on the executive. It could also concern the direct cost of implementing the policy change which can be uncertain (normalized to zero in our case).

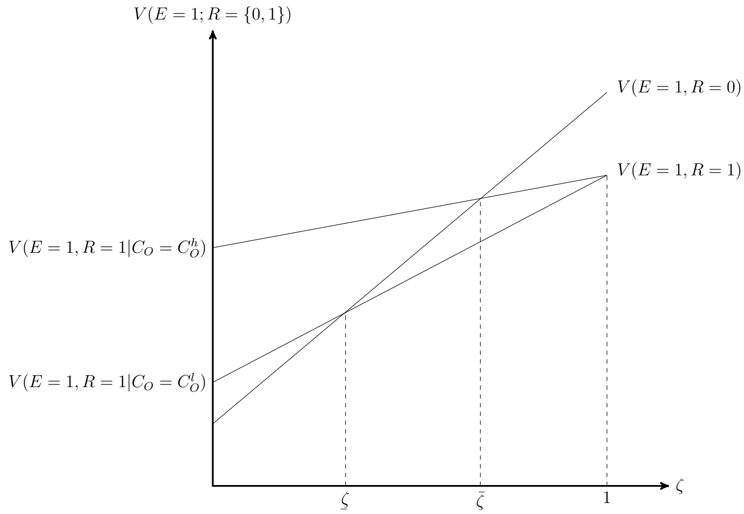

Figure 1: Post-experiment choice to reform.

where $i = \{l, h\}$. When $V_I\,(R = 1; E = 1; C_O = C_O^i) > V_I\,(R = 0; E = 1)$, the incumbent will find it profitable to implement a policy change following the experiment. We can state the first result, which defines a threshold values for the degree of political competition.

Proposition 1. *There exists a value $\underline{\zeta} \in [0; 1]$ and $\overline{\zeta} \in [0; 1]$ with $\overline{\zeta} > \underline{\zeta}$ such that (i) when $\zeta < \underline{\zeta}$ the incumbent will always implement a policy change following an experiment, (ii) when $\zeta > \overline{\zeta}$ the incumbent will never implement a policy change following an experiment, and (iii) when $\underline{\zeta} < \zeta < \overline{\zeta}$ the incumbent will implement a policy change only if $C_O = C_O^h$. We can show , $\underline{\zeta} = C_O^l/(C_O^l + C_I)$ and $\overline{\zeta} = C_O^h/(C_O^h + C_I)$.*

Proof. See Appendix. □

The value functions $V_I\,(R = 0; E = 1)$ and $V_I\,(R = 1; E = 1; C_O = C_O^i)$ for $i = \{l, h\}$ are represented on Figure 1. Note that when $\zeta = 1$, both value functions of implementing the policy change $V_I\,(R = 1; E = 1; C_O = C_O^i)$ are equal since the probability of being in power the next period is equal to one. And this value is clearly lower than the value function of not implementing the policy change since the position of the incumbent is worse when there exist some constraints on power. Also, when $\zeta \neq 1$ the value function of implementing the policy change when C_O is high is always higher than when C_O is low. Finally, when $\zeta = 0$ the incumbent has a probability one of being part of the opposition next period and the value of implementing the policy change after the experiment is always higher than the value of not implementing the policy change. Following this logic the result of Proposition 1 is quite intuitive. When ζ is high ($\zeta > \overline{\zeta}$), it is never profitable for the incumbent to implement a policy change. Indeed the incumbent has a very high probability of being in power the next period and constraining power will thus very likely lead to a utility loss. When ζ is low ($\zeta < \underline{\zeta}$), then the incumbent always has an incentive to pursue the reform following an experiment for any value of C_0. Since the probability the incumbent remains in power is low, he will likely be in the opposition in the next period and constraining power will lead to a utility gain in expectation. For intermediate values of ζ (i.e. $\underline{\zeta} < \zeta < \overline{\zeta}$), the probability of losing power is quite high and the incumbent will implement the reform if and only if the the cost (utility loss if he remains in power) is not too high compared to the benefit (utility gain if he is in opposition the next period). This is the case if and only if the gain C_O is high enough ($C_O = C_O^h$). Here appears very clearly one of the benefits of running an experiment: it allows the incumbent to implement a policy change with a better knowledge of the cost and benefit implied by the policy change. We now turn to the decision of running an experiment.

3.4 The pre-experiment decision

Here we analyze the value of running an experiment for the incumbent . We start by presenting the value functions of the incumbent under the four possible combinations in $P(E;R)$. First, when not running an experiment ($E = 0$) and not making the policy change ($R = 0$), we have:

$$V_I\left(R = 0; E = 0\right) = R^I + \beta\left[\zeta R^I\right] \tag{3}$$

The incumbent is re-elected for a second term with probability ζ and enjoys the rent from being in office the next period if re-elected. With probability $1 - \zeta$ the incumbent will be in the opposition group for the next period and his utility is normalized to zero when there are no constraints on power.

Second, when implementing the policy change ($R = 1$) without running an experiment beforehand ($E = O$), we have:

$$V_I\left(R = 1; E = 0\right) = R^I + \beta\left[\zeta\left(R^I - P_R C_I\right) + \left(1 - \zeta\right) P_R E(C_O)\right] \tag{4}$$

$E(C_O)$ is the expected value of C_O, equal to $hC_O^h + (1-h)C_O^l$. As before, the incumbent is re-elected for a second term with probability ζ, but when pursuing the policy change, the rent is reduced by C_I if the reform is effective, which occurs with probability P_R. The incumbent is not re-elected with probability $(1 - \zeta)$, however, in this case he enjoys an additional expected utility $E(C_O)$ of being in the opposition group if the reform is effective (again, with probability P_R). The trade-off appears clearly here. The benefit of implementing a reform to constrain power accrues if the incumbent is in the opposition in the next period, as he would gain from tying the hands of his opponent. On the other hand, the reform is costly for the incumbent if he is re-elected in the next period as it would effectively tie his own hands.

It is also clear from equation (4) that the value of running an experiment will depend on the level of political competition, ζ, given that it affects the decision to implement the policy change after the experiment and will ultimately affect the value of running an experiment as a result. We distinguish between several cases depending on the value of ζ. The first case is when $\zeta < \underline{\zeta}$. In this case, the incumbent always implements the policy change ($R = 1$) after running an experiment and the value function of running an experiment is:

$$V_I \left(E = 1, \zeta < \underline{\zeta}\right) = R^I + \beta \left[\zeta \left(R^I - P_E C_I\right) + (1 - \zeta) P_E E(C_O)\right] - c^e \tag{5}$$

Here, when $\zeta < \underline{\zeta}$ the incumbent undertakes the policy change whatever C_O is revealed to be. The only difference with equation (4) is that the policy change is more likely to be effective in constraining power (i.e., $P_E > P_R$).

When $\underline{\zeta} < \zeta < \overline{\zeta}$, the incumbent undertakes a policy change only if the experiment reveals that the benefit to do so is sufficiently high (i.e $C_O = C_O^h$). In this case the value function of the incumbent is:

$$V_I \left(E = 1, \underline{\zeta} < \zeta < \overline{\zeta}\right) = R^I + \beta \left[\zeta \left(R^I - P_E h C_I\right) + (1 - \zeta) P_E h C_O^h\right] - c^e \tag{6}$$

The only difference with (5) is that when $\underline{\zeta} < \zeta < \overline{\zeta}$, the incumbent undertakes a policy change if and only if $C_O = C_O^h$ which occurs with probability h. Therefore, the incumbent incurs the cost C_I and benefit C_O^h of the reform with probability h.

When $\zeta > \overline{\zeta}$, the incumbent never implements the policy change following the experiment whatever the value of C_O. The value function of the incumbent is thus:

$$V_I \left(E = 1, \zeta > \overline{\zeta}\right) = R^I + \beta \left[\zeta R^I\right] - c^e, \tag{7}$$

which is always lower than $V_I (R = 0; E = 0) = R^I + \beta \left[\zeta R^I\right]$ since the cost of running an experiment is strictly positive. When $\zeta > \overline{\zeta}$, the incumbent will never have an incentive to run an experiment.

We now turn to the analysis of the decision to run an experiment. We first show that there exists a critical value for the degree of political competition above which, when no experiment has been run, the incumbent prefers not to implement a policy change. Second, we show that if the cost of running an experiment is not too low, there exists a degree of competition below which the incumbent prefers to run an experiment rather not implementing the policy change at all. We finally show that when there is no cost of running an experiment, the incumbent

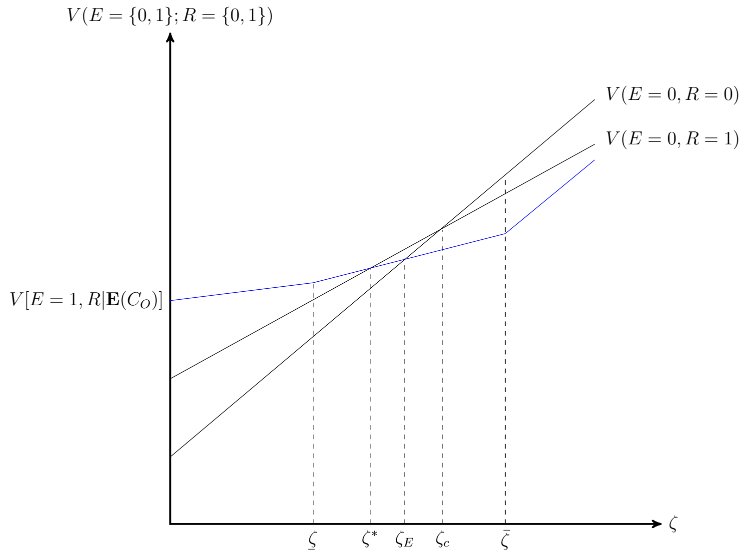

Figure 2: Pre-experiment choice for an intermediate cost of the experiment.

always has an incentive to run an experiment prior to eventually reforming. From this result we discuss the different equilibrium configurations depending on the cost of running an experiment.

Proposition 2. *There exists a unique value $\zeta_c \in \left[\underline{\zeta}, \overline{\zeta}\right]$ such that $V_I\left(R=0;E=0\right) = V_I\left(R=1;E=0\right)$. When not having run an experiment ($E=0$), below this threshold ($\zeta < \zeta_c$) the incumbent finds it optimal to implement a policy change (that is, $V_I\left(R=0;E=0\right) < V_I\left(R=1;E=0\right)$). Above this threshold ($\zeta > \zeta_c$), the incumbent does not find it optimal to implement a policy change (that is $V_I\left(R=0;E=0\right) > V_I\left(R=1;E=0\right)$). We can show $\zeta_c = E(C_O)/(E(C_O)+C_I)$.*

Proof. See Appendix. □

The intuition of the result is best seen graphically, shown in Figure 2. If the incumbent does not have the option to run an experiment, he implements the policy change only if the probability of losing power is sufficiently high (low ζ). This is because he will be in the opposition group with a high probability in period 2 and thus puts a high weight on the opposition's utility, which increases the incentive to implement a constraint on power. We now analyze the decision to run an experiment.

Proposition 3. *When $\beta P_E E(C_O) > c^e$ there exists a unique $\zeta_E \in \left[0, \overline{\zeta}\right]$ such that $V_I\left(E=1\right) = V_I\left(R=0;E=0\right)$. Below this threshold, (when $\zeta < \zeta_E$) the incumbent finds it optimal to run an experiment rather than not running an experiment and not reforming (that is, $V_I\left(E=1\right) > V_I\left(R=0;E=0\right)$). Above this threshold ($\zeta < \zeta_E$) the incumbent prefer not to* *reform (and not run an experiment) rather than running an experiment (that is $V_I(E=1) < V_I(R=0;E=0)$). We can show that when $\zeta_E < \underline{\zeta}$ then $\zeta_E = (E(C_O) - (c^e/\beta P_E))/(E(C_O) + C_I)$ and when $\zeta_E > \underline{\zeta}$, then $\zeta_E = (C_O^h - (c^e/\beta P_E))/(C_O^h + C_I)$*

Proof. See Appendix. □

Figure 2 provides a graphical representation of this result.

The following assumption simplifies the analysis that follows.

Assumption $\beta\left[\zeta\left(R^I - P_E h C_I - P_E h C_O^h\right)\right] < \beta\left[\zeta R^I - P_R C_I - P_R E(C_O)\right]$ which is satisfied if $P_E h C_I + P_E h C_O^h > P_R C_I + P_R E(C_O)$. A sufficient (but not necessary) condition would be that $P_E h > P_R$ since $C_O^h > E(C_O)$.

This assumption ensures that the the slope of $V_I(E=1)$ is steeper than $V_I(R=1;E=0)$ in the $[\underline{\zeta};\overline{\zeta}]$ interval. It also ensures that both value functions intersect only once in the $[0;\overline{\zeta}]$ interval if an intersection exists, as it ensures that $V_I(E=1)$ is steeper than $V_I(R=1;E=0)$ in the $[0;\underline{\zeta}]$ interval (see Figure 2). The equilibria that we analyze in the three cases below are differentiated by the cost of running an experiment. As a benchmark, we first deal with with the extreme case for which there is no cost of running an experiment ($c^e = 0$).

Proposition 4. *When there is no cost of running an experiment ($c^e = 0$), the incumbent always has an incentive to run an experiment prior to eventually implementing a policy change. More precisely we have $V_I(E=1) \geq \max\{V_I(R=1;E=0);V_I(R=0;E=0)\}$.*

Proof. See appendix □

This intuitive result is represented in Figure 3. When $\zeta \geq \overline{\zeta}$, it is never profitable for the incumbent to implement a policy change after running an experiment, since as the cost of running an experiment is zero, $V_I(E=1, \zeta \geq \overline{\zeta}) = V_I(R=0;E=0)$. When $\zeta < \overline{\zeta}$ the incumbent always finds it profitable to implement the policy change following the experiment when C_O is low. When running an experiment costs nothing, he always has an incentive to run an experiment when $\zeta < \overline{\zeta}$ to find out about C_O.

We now provide a simple graphical analysis of the three possible equilibrium configurations we can have in this simple model. First note that an increase in c^e simply shifts $V_I(E=1)$ downward without any impact on the other value functions.

Case 1: Low-cost experiments. There exist some low values for c^e such that the incumbent always runs an experiment before deciding about a policy change. Recall that when $c^e = 0$ the incumbent always runs an experiment for $\zeta < \overline{\zeta}$. The reasoning should hold

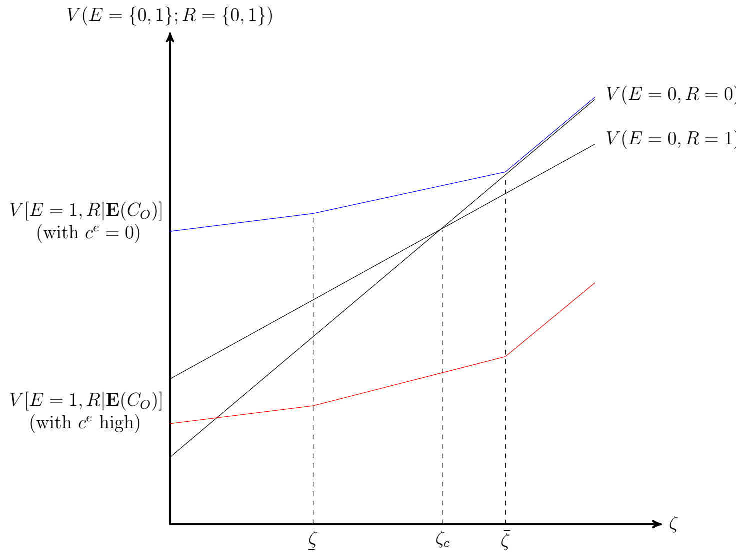

Figure 3: Pre-experiment choice for different experiment cost parameterse.

for some positive but sufficiently low values of $c^e > 0$ since an increase in c^e simply shifts $V_I(E=1)$ downward (see Figure 3). In this case, we can see that $V_I(E=1)$ is always higher that $V_I(R=1; E=0)$ for any ζ lower than ζ_c, the threshold value of political competition above which the incumbent prefers to not reform (without experiment) rather than reforming without running an experiment. In this case, the cost of running an experiment is sufficiently low to make it worthwhile for the incumbent given the level of political competition and the expected benefit of a reform. For low value of ζ ($\zeta < \underline{\zeta}$), the experiment is always followed by a policy change. The experiment allows a better design of the policy and increases the probability the reform will be effective ($P_E > P_R$). For intermediate values of ζ ($\underline{\zeta} < \zeta < \overline{\zeta}$) the experiment is not always followed by a policy change. Here the experiment also allows to avoid implementing a policy which is ex-post non-profitable (i.e when $C_O = C_O^l$) and the experiment is not necessarily followed by a policy change.

Case 2: Intermediate-cost experiments. There exist some intermediate values for c^e such that the incumbent finds it profitable to run an experiment prior to deciding about a policy change only when ζ is sufficiently low, but prefers to implement the policy change without running an experiment for some intermediate values of ζ. This case is represented in Figure 2. As c^e increases and $V_I(E=1)$ shifts downward, at some point it should intersect $V_I(R=1; E=0)$ at a value ζ^* which is lower than ζ_c (which should necessarily occur when $\zeta_E < \zeta_c$). When $\zeta < \zeta^*$, the incumbent will run an experiment prior to reforming and when $\zeta_c > \zeta > \zeta^*$ the incumbent will introduce the policy change without running an experiment (and of course when $\zeta > \zeta_c$ the incumbent does not run an experiment and does not implement a policy change). In the case represented in Figure 2, $\zeta^* > \underline{\zeta}$ and some of the experiments will not be followed by a policy change as a result.[7]

Case 3: High-cost experiments. There exist some high values for c^e such that the incumbent never finds it optimal to run an experiment. This case is represented in the Figure 3. To see this, note that when $\zeta = 0$ we always have that $V_I\left(R=1; E=0\right) > V_I\left(R=0; E=0\right)$ and that $V_I\left(E=1, \zeta < \underline{\zeta}\right) < V_I\left(R=1; E=0\right)$ if and only if $c^e > (P_E - P_R)E(C_O)$. In other words, when c^e is too high the intercept of $V_I(E=1)$ is lower than $V_I\left(R=1; E=0\right)$ and it will never be profitable to run an experiment even when the incumbent is certain to lose power ($\zeta = 0$).

3.5 From the theory to the empirics.

We can now derive several empirical predictions from the model which can be tested empirically.

Hypothesis 1 The probability to observe an experiment increases with the degree of political competition.

The is the first clear prediction of the model. In case 1, when the probability of losing power is sufficiently high, the incumbent will run an experiment and eventually implement a policy change. In case 2, when ζ decreases, the incumbent first finds optimal to implement a policy change without running an experiment and then, for lower values of ζ, the incumbent find it optimal to run an experiment. As a result, we should observe a strong correlation between the probability of observing an experiment in a given year/area and proxies for political competition.

Hypothesis 2 The impact of political competition on the probability to observe an experiment is greater in politically polarized society.

This is the second prediction of our model. For instance, in highly polarized societies the benefits the opponent gets from constraint on power (the $E(C_O)$ parameter) should be high. Conversely, in more uniform societies the political turnover should be less costly for the politician given that policy platforms are closer and politicians should not use discretionary power to strongly favor one group. Given this, when $E(C_O)$ is lower the value function $V_I\left(E=1, \zeta < \underline{\zeta}\right)$ shifts downward and we end up in the case 3 configuration in which the probability to observe an experiment is not related at all to the degree of competition ζ. On the other hand, when $E(C_O)$ becomes larger, the value function $V_I\left(E=1, \zeta < \underline{\zeta}\right)$ has a

[7] Note that when c^e is sufficiently low (possibly equal to zero), then $\zeta_E > \zeta_c$ (in case $c^e = 0$, then $\zeta_E = \overline{\zeta} > \zeta_c$, see Figure 3) and we are in case 1 in which there is always an experiment prior to reforming.

higher intercept $[V_I\left(E=1,\zeta<\underline{\zeta}\right)$ evaluated at $\zeta=0]$, which should end in case 2 or 1 in which the probability to observe an experiment react more to competition compared to case 3.

We can also identify a third empirical prediction which we can directly infer from the logic of our model: political competition should increase the share of experiments in which the government is involved. This is not a direct prediction of the model since we have not modeled the partner choice dimension. However, in our model the incumbent government has a direct incentive to implement a policy change and to run an experiment to learn about the design and effect of the policy change. Political competition affects those incentives. One can think that many other actors (NGOs, academics and others) run experiments for very different motives, others than political competition. As a result, it is quite natural to consider that the proportion of experiment in which the government is involved should increase with political competition.

The focus of our formal model has been on a "tying-the-hands" incentive for the policy reform, which is strengthened by running an experiment. Thus, the most empirically relevant experiments for testing the specific mechanism that we have highlighted in the model are those that investigate the accountability of the public sector. However, we note that any kind of experiment can also tie the hands of a government if the results of the experiment demonstrate how specific policies can be welfare-enhancing for the population. In our empirical investigation, we examine the probability of running an experiment of any kind, and we also categorize the experiments according to the type of policy that they investigate.

4 Data

In order to test the hypotheses that we have developed in the formal model, we generated two original datasets. The first dataset contains 289 RCT experiments conducted in 43 countries between 1995 and 2015. We followed four main selection and coding criteria. First, we identified the Abdul Latif Jameel Poverty Action Lab (J-PAL) as the single source for our RCT sample. J-PAL is currently the leading research and policy network conducting field experiments in international development. This criterion allows us to control for various factors that might affect the supply of experiments such as personal contacts, the reputational vetting of organizational partners or a high likelihood of scale-up, and keep the quality of design and scientific rigor constant. Second, we selected three policy areas – political economy and governance, health, and education – that are the most salient in international development. Third, we chose to code all the RCTs in these three areas that were recorded online in the J-PAL study repository as of August 2016. Therefore, our time cut-off point is currently 2015.

Fourth, we excluded the few experiments that were not randomized control trials or that investigated historical rather than contemporary data retrospectively. The cross-national dataset includes a wide range of variables, such as the implementation partners, the number of treatments, the number of results, the type of evaluation, the conditional or unconditional nature of the incentive used during the intervention, and the type of treatment transfers. We have additionally merged the J-PAL dataset that we constructed with publicly available macro-level data that capture the degree of political competition, our primary explanatory variable, as well as macro-economic, socio-economic, and demographic control variables.

Our second dataset considers the subset of 166 J-PAL experiments taking place in India between 1995 and 2015, as well as a series of state-level variables meant to capture the types of political institutions, macroeconomic controls, and mitigate omitted variable bias. We selected India as country of analysis for two reasons. First, it is the leading host of RCT experiments in the field of international development. Second, despite being pilots, many RCTs conducted in Indian states, involve a large enough subject pool to attract concrete reactions from politicians and bureaucrats, independently of the decision to scale up or not. Following the same coding protocol, we coded the individual characteristics of the experiments, and collected additional political economic variables across 29 Indian states and 7 Union territories based on primary sources.

4.1 Categorizing the RCT's

We then generated six categories concerning the function of the experiments in order to match the parameters of the theoretical model that reflect the policy-makers' present and future costs and benefits associated with hosting an RCT evaluation. The categories that we considered are (i) *policy learning*, (ii) *electoral learning*, (iii) *political accountability*, (iv) *bureaucratic accountability*, (v) *decentralization*, and (vi) *minority inclusion*. We allowed for experiments to fit into multiple categories. *Policy learning* contains experiments that entail precise tests of the impact of a policy on a technical outcome of interest. For instance, a celebrated evaluation assessed the impact of deworming treatments for children on school participation in Kenya (Miguel and Kremer, 2004). *Electoral learning* consists of experiments whereby researchers, bureaucrats or politicians test the most efficient ways to appeal to voters, fundraise, increase voter turnout and electoral participation, or shape electoral outcomes. The RCT conducted in Benin that tested the effect of clientelistic messages on voter behavior offers once such example (Wantchekon, 2003). *Political accountability* experiments comprise RCTs that directly test policy tools that could increase the responsiveness of elected public officials to citizens, voters and public service users. An RCT conducted in Mexico evaluated, for instance, how information dissemination about politicians' corruption and budget expenditure affected election participation. *Bureaucratic accountability* is a related category of experiments that test the best ways to reduce the discretionary decision-making of bureaucrats rather than politicians. The large scale Indian RCT testing the effect of biometric smart cards on reducing bureaucratic leakage and corruption in social benefit disbursement fits into this category (Muralidharan et al., 2016). *Decentralization* refers to the family of experiments that test the effect of community participation in co-producing, managing or monitoring public services traditionally run by the state. For example, a field experiment conducted in Sierra Leone estimated the impact of village development committees on local infrastructure investment. Last, *minority inclusion* experiments include the RCTs of our dataset that attempt to find the most effective ways to include gender, racial, ethnic, linguistic minorities, or scheduled caste and tribes traditionally marginalized in formal or informal decision-making institutions, or that assess the impact of specific policy tools of minority representation on the general performance of the members of minority groups. One such experiment conducted in India evaluated the impact of women political leadership at the village level on the social aspirations of children and parents. The main theoretical difference among these six categories consists in the different configurations of RCT costs and benefits for incumbent policy-makers. We argue that electoral learning experiments (category ii) entail high benefits and low costs, whereas accountability and decentralization RCT (categories iii, iv, and v) imply high costs for incumbents, but potentially high benefits if they lose power in future elections. Somewhere in the middle of the spectrum stand the experiments related to policy learning and minority inclusion (categories i and vi) whose benefits and costs vary according to a specific policy's distance from the ideal point of decision makers. A detailed description of these categories, as well as reporting of an inter-coder reliability exercise are reported in the Online Appendix.

Figure 5.3 shows a few trends in the number of RCT experiments conducted for the two datasets and how many of them were in India (the lighter grey line). In the top left panel, we show the number of RCT's by year. In the top right panel, we show the number of RCT's which included the government as a partner. The bottom panels show two of our most important categories: on the left, the number of RCTs that were categorized as "policy learning" and on the right, the number that were classified as *public sector accountability* (which includes all three sub-categories *political accountability*, *bureaucratic accountability*, and *decentralization*).

A premise of the formal model from the last section is that the government is involved with the experiments. However, not all experiments are run with the government as an explicit partner. We have thus coded whether or not the government was a partner in

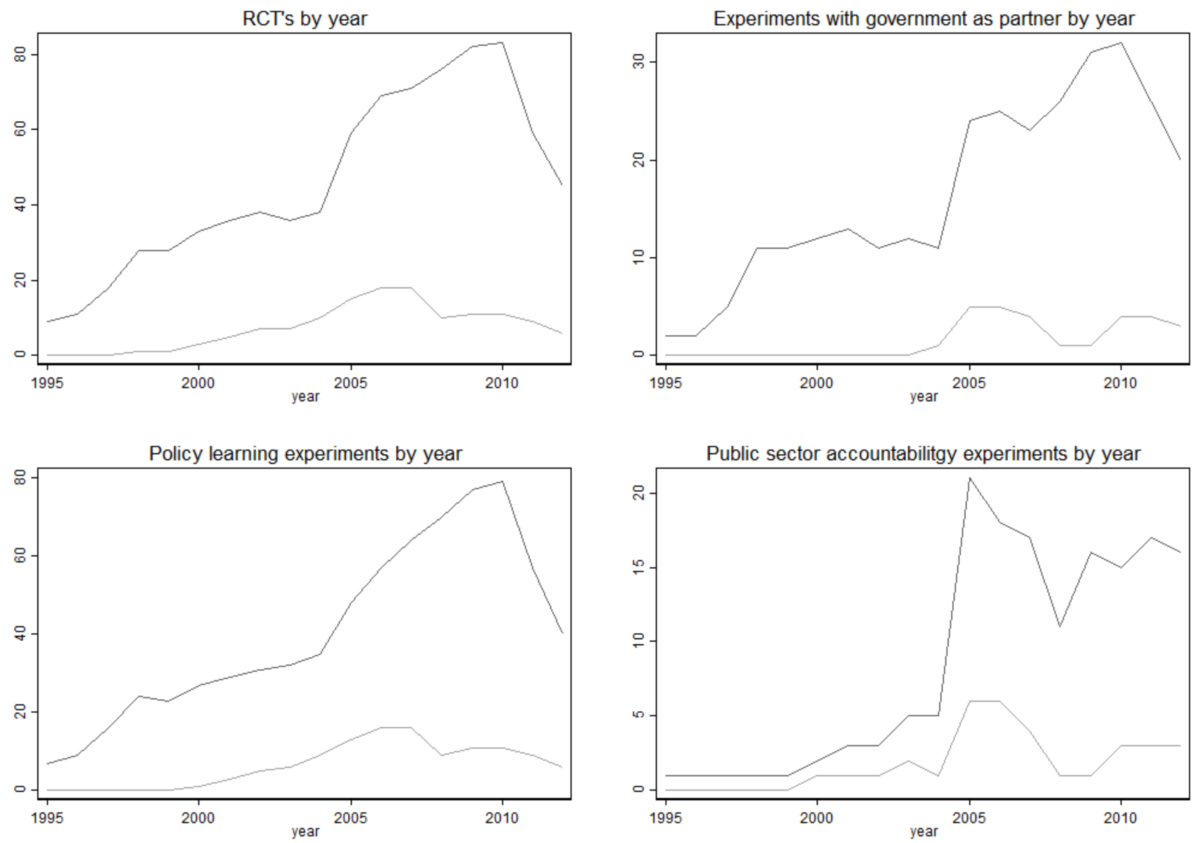

Figure 4: The evolution of RCT's since 1995. The top line in each panel shows the number of RCT's overall, while the bottom line in each panel shows the number of RCT's taking place in India. Note that we have cut the figure at 2012, as RCT's are recorded on the JPAL repository only with a significant lag.

the experiment. Table 1 tabulates experiments by category and by whether or not the government was an explicit partner in executing the experiment. In accord with our model, we will later demonstrate that political competition can explain the occurrence of RCTs that have the government as a partner, but cannot explain the occurrence of RCTs that do not have the government as a partner.

4.2 Political competition and additional controls

For the cross-national analysis, our primary explanatory variable is the index of the competitiveness of executive recruitment from the Polity IV project, where competitiveness refers to "the extent that prevailing modes of advancement gives subordinates equal opportunities to become superordinates" (Marshall et al., 2016). The index is increasing in the degree of competition, and we have labeled this variable as *political competition*. Importantly, the sub-index of the Polity IV data that we employ as a measure of political competition is not

Table 1: Experiments with government involvement

	Indian data			Cross-national data		
Experiment type	Gov't	No gov't	Share gov't	Gov't	No gov't	Share gov't
Any experiment	22	75	0.23	152	160	0.49
Policy learning	22	67	0.25	138	140	0.50
Electoral learning	0	12	0.00	26	24	0.52
Public sector accountability	17	10	0.63	72	30	0.71
Political accountability	4	7	0.36	23	25	0.48
Bureaucratic accountability	13	3	0.81	48	2	0.96
Decentralization	0	2	0.00	37	10	0.79
Minority inclusion	1	18	0.05	17	8	0.68

Notes: Binary variables take value 1 if an experiment happened in a country/year for the various experiment types. Note that experiments can be coded to have more than one type.

related to the degree of executive constraint. We have also considered a range of alternative cross-country proxies for the degree of political competition, which we present in the Online Appendix.[8]

To measure political competition in India we considered the ruling margin of seats in state parliaments. The *margin of seats* variable equals the seats held by the ruling government minus the seats held by the first runner-up, divided by the total number of seats in the state parliament. A smaller seat margin signifies a more robust level of political competition.[9]

We have tried to include the same battery of control variables for both the cross-country regressions and the Indian state-level regressions. Besides the political competition variables, additional controls include income per capita, population, state capacity proxies, rates of urbanization, indicators for elections, and proxies for political polarization. Summary statistics for the two samples can be found in Online Appendix Tables A1 and A2.

[8]We have considered the competitiveness of participation index from the Polity IV project and the measure proposed by Vanhanen (2016). The competitiveness of participation refers to "the extent to which alternative preferences for policy and leadership can be pursued in the political arena" (Marshall et al., 2016). Vanhanen (2016) is the closest to the measure that we employ for the Indian panel. It is calculated to portray the electoral success of smaller parties in parliamentary and/or presidential elections.

[9]We have additionally computed a Herfindahl-Hirschman index of seat concentration. Results using the HH index are presented in the Online Appendix.

5 Empirical investigation

5.1 Empirical methodology and initial results

This section presents the results of our empirical investigation into the political economy of randomized controlled trials (RCT's). Using proxies for different aspects of political competition, we examine the extent to which local political conditions impact the likelihood that a locality hosts an RCT. To do so, we consider panel regressions with a binary dependent variable, $RCT_{i,t}$, which takes value 1 if a program evaluation was launched in district i during year t and 0 otherwise. As described above, we have also coded the type of RCT, in terms of the objective of the experiment. Our baseline regression analysis thus also considers binary dependent variables defined according to the RCT type, as well as continuous dependent variable which is the share of RCTs of a certain type. Formally, we use fixed effects linear regression models to estimate the following:

$$RCT_{i,t} = \alpha PolComp_{i,t} + \beta X_{i,t} + \gamma_i + \delta_t + u_{i,t}, \tag{8}$$

where $PolComp_{i,t}$ is a generic label for the different measures of political competition that we will consider, X is a vector of control variables that will include a per capita income measure and the jurisdiction's population in the baseline specifications, the γ_i's denote a full set of entity dummies that capture any time-invariant entity characteristics that affect the likelihood of a program evaluation, and the δ_t's denote a full set of period dummies that capture common shocks to the likelihood of a program evaluation. The error term $u_{i,t}$ captures all other factors not correlated with our controls which may also explain democratic switches, with $E(u_{i,t}) = 0$ for all i and t. All estimations are conducted with standard errors that are clustered at the entity level.

We begin by presenting some results about whether political competition can explain experiments that were ran in partnership with the government better than those that were not. Note that for this analysis, we do not distinguish between the type of experiment, but are instead interested in investigating whether the government is involved with carrying out the experiment. Columns 1 – 3 of Table 2 show some results for the Indian state-level panel. In the first column the binary dependent variable takes value one if the government hosts an experiment on which it was a partner. The result shows that greater political competition increases the likelihood that a government participates in an RCT. In the second column, the binary dependent variable takes value one if the government hosts an experiment in which it was not a partner. Here, we see that political competition cannot explain whether or not a government hosts an experiment in which it did not participate. Finally, in the third column

Table 2: Experiments with governmental involvement and political competition

	Fixed effects OLS regressions					
	Indian data			Cross national data		
	Government binary	No gov't binary	Government share	Government binary	No gov't binary	Government share
	(1)	(2)	(3)	(4)	(5)	(6)
Margin of seats	-0.060*	0.057	-0.046*			
	(0.03)	(0.04)	(0.02)			
p_xrcomp				0.035**	-0.001	0.034**
				0.02)	(0.02)	(0.02)
ln(income)	-0.038	0.084	-0.004	-0.012	-0.009	-0.010
	(0.07)	(0.17)	(0.04)	(0.03)	(0.02)	(0.02)
ln(population)	-0.142	0.057	-0.090	0.024	0.097	0.015
	(0.19)	(0.30)	(0.16)	(0.06)	(0.07)	(0.05)
District & year FE's	yes	yes	yes	yes	yes	yes
N	553	553	553	2607	2607	2607
Districts / Countries	30	30	30	160	160	160
N. experiments	x	x	x	143	140	283

Notes: Robust standard errors clustered by entity are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

the dependent variable is a continuous variable that is the share of experiments hosted by the government for which the government is a partner or a principal investigator. The statistically significant negative coefficient on the ruling seat margin confirms that political competition is an important factor for explaining when governments participate in RCTs. Columns 4 – 6 of Table 2 show the analogous regressions for the cross-national panel. Results are similar for this sample of countries using the competitiveness of executive recruitment measure from the Polity IV project as our measure of political competition.

We will now proceed to investigate how political competition affects the types of experiments differentially. We will first present the results from regressions that use the cross-national data before zooming in on the state-level Indian data. Tables 3 – 5 present the baseline results for the cross-country data, using p_xrcomp (competitiveness of executive recruitment) as the proxy for political competition. Table 6 then investigates whether the effect of political competition is amplified by political polarization, as our formal theory predicts. Tables 7 – 9 then present the baseline results for the Indian data, using the margin of seats in district parliaments as the proxy for political competition. We have conducted many robustness tests on our main results that are presented in the Online Appendix, to which we will refer interested readers where appropriate.

Table 3: Experiments and competitiveness of executive recruitment (p_xrcomp)

Panel A	Binary DV: Fixed effects LPM regressions				
	All RCTs	Policy learning	Electoral learning	Account -ability	Minority inclusion
	(1)	(2)	(3)	(4)	(5)
political competition	0.033	0.024	0.006	0.048**	0.006
	(0.02)	(0.02)	(0.00)	(0.02)	(0.00)
ln(gdppc)	-0.021	-0.006	-0.002	0.005	0.012
	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)
ln(population)	0.120	0.073	0.029	0.110*	0.027
	(0.10)	(0.09)	(0.03)	(0.06)	(0.02)
Panel B		**Shares DV: Fixed effects OLS regressions**			
		Policy learning	Electoral learning	Account -ability	Minority inclusion
		(1)	(2)	(3)	(4)
political competition		0.025	0.006	0.041**	0.001
		(0.02)	(0.00)	(0.02)	(0.00)
ln(gdppc)		-0.007	-0.000	-0.018	0.000
		(0.03)	(0.00)	(0.02)	(0.00)
ln(population)		0.068	0.029	0.107*	0.010
		(0.09)	(0.03)	(0.06)	(0.01)
District & year FE's	yes	yes	yes	yes	yes
N	2607	2607	2607	2607	2607
Countries	160	160	160	160	160
N. experiments	283	253	45	94	23

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

5.2 Results with cross-national data

In this sub-section, we begin our empirical investigation at the cross-country level. From 1995 until 2015, our database covers 289 program evaluations which took place in 43 countries. In Table 3, we have organized the results from the cross-national panel according to the type of experiment. In Panel A we consider binary dependent variables, while in Panel B we consider the shares dependent variables. We consider whether or not there was an experiment at all (in column A1), whether or not there was a *policy learning* experiment (column A2), whether or not there was an *electoral learning* experiment (column A3), whether or not there was a *public sector accountability* experiment (column A4), and whether or not there was a *minority inclusion* experiment (column A5). In Panel B, we consider the analogous

Table 4: Experiments and competitiveness of executive recruitment (p_xrcomp)

	Binary dependent variables							
	Fixed effects logit regressions				Random effects logit regressions			
	All RCTs	Policy learning	Electoral learning	Account -ability	All RCTs	Policy learning	Electoral learning	Account -ability
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
political competition	0.345	0.306	0.373	1.159***	0.759***	0.709***	1.738**	1.324***
	(0.25)	(0.25)	(1.14)	(0.38)	(0.23)	(0.23)	(0.80)	(0.32)
ln(gdppc)	2.210**	2.726***	0.313	4.642***	-0.735***	-0.520**	-0.431	-1.296***
	(0.87)	(0.89)	(1.59)	(1.50)	(0.24)	(0.24)	(0.45)	(0.32)
ln(population)	3.135	2.224	11.724	0.519	1.679***	1.706***	1.981***	1.633***
	(3.15)	(3.29)	(7.91)	(4.67)	(0.31)	(0.32)	(0.66)	(0.35)
Country FE's	yes	yes	yes	yes	no	no	no	no
Year FE's	yes	yes	yes	yes	yes	yes	yes	yes
N	690	644	148	304	2607	2607	2607	2607
Countries	42	39	9	19	160	160	160	160
N. experiments	266	236	45	94	283	253	45	94

Notes: *** / ** / * represent significance at the 1 / 5 / 10 percent level.

categories, as the share of experiments that took place (columns B1 – B4). To measure political competition, in the baseline analysis we employ the measure of competitiveness of executive recruitment from the Polity IV project (p_xrcomp). The results of Table 3 reveal that political competition is not statistically significantly correlated with the likelihood that a country hosts a RCT. However, there is a positive and statistically significant correlation between political competition and the likelihood of a public sector accountability experiment. In other words, an increase in the degree of electoral competition makes hosting an accountability RCT more likely. There is no statistically significant relationship between political competition and the other kinds of experiments.[10]

In Table 4, we present binary dependent variable model results obtained with non-linear estimators. In columns 1 – 4, we present results from a fixed-effects logit model, while in columns 5 – 8, we present results from a random-effects logit model. We note that neither of the estimators are perfectly suitable for our purposes. The fixed-effects estimator drops countries for which there is no variation in the dependent variable, which dramatically reduces our sample sizes. Nonetheless, our main result from Table 3 is verified with the fixed-effects logit estimator: political competition affects the likelihood of holding a public sector accountability experiment, but does not affect the likelihood of the other kinds of

[10]Table A3 of the online appendix demonstrates that the result is robust to estimation over a sample that drops the advanced industrialized democracies. Table A4 of the online appendix considers more fine-grained categories within the *public sector accountability* category. Specifically, we consider the categories for *political accountability*, *bureaucratic accountability*, and *decentralization*. The result appears to be driven by the bureaucratic accountability experiments.

Table 5: Accountability experiments and competitiveness of executive recruitment (p_xrcomp)

	Share accountability experiments DV: Fixed effects OLS regressions						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
political competition	0.041**	0.116***	0.041**	0.113**	0.059**	0.041**	0.116***
	(0.02)	(0.04)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)
ln(gdppc)			-0.018	-0.037	-0.038	-0.016	-0.060
			(0.02)	(0.03)	(0.03)	(0.02)	(0.05)
ln(population)			0.107*	0.282***	0.159	0.093	0.104
			(0.06)	(0.11)	(0.10)	(0.06)	(0.17)
parliamentary election			-0.006				-0.006
			(0.00)				(0.01)
tax revenues				-0.000			-0.003*
				(0.00)			(0.00)
education expenditures					0.016*		0.029
					(0.01)		(0.02)
urbanization						0.004	0.006
						(0.00)	(0.01)
Country & year FE's	yes	yes	yes	yes	yes	yes	yes
N	2607	1037	2607	1551	1533	2607	1037
Countries	160	120	160	136	148	160	120
N. experiments	94	50	94	68	65	94	50

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

experiments, nor of the likelihood of an experiment in general. The random-effects estimator, on the other hand, does not isolate within-country variation and thus leaves open the door for omitted variable bias. Furthermore, the estimates seem to be biased upwards, which may be artificially increasing the ratio of the estimated coefficient to its standard error. We see from columns 5 – 8 that the random-effects logit model estimates that an increase in political competition increases the likelihood of each type of RCT.[11]

In Table 5, we present a basic robustness analysis of the results with the share of governmental accountability estimation (from column B3 of Table 3) by including some additional time-varying controls whose omission could cause concern. First, column 1 presents the raw bivariate correlation, which we see is identical to that of our baseline specification. As we lose many observations in the model with the full set of regressions (column 7), in column 2 we show the simple bivariate regression over the equivalent sample from column 7. Adding a dummy variable for whether or not a parliamentary election was held in the country-year does not affect the estimation (column 3). Column 4 controls for tax revenues as a fraction

[11] The logit estimator does not converge for the minority inclusion dependent variable, so we omit models that feature dependent variables based on this category from the table.

of GDP (a proxy for state capacity that may attract experimenters), column 5 controls for public education expenditures as a fraction of GDP (a proxy for the capacity of citizens), column 6 controls for the rate of urbanization (a proxy for the supply side of experiments, since most are held in villages), and column 7 adds all four of the additional controls at once to the panel regression equation. The result is qualitatively robust to these additional time-varying controls. The estimated effect of political competition remains positive and statistically significant. While the point estimate is sensitive to the addition of the state capacity proxy (columns 4 and 7), we note that the change in the estimate is clearly due to the difference in the sample when this control is included.[12]

5.3 Testing for the effect of political polarization

We next examine our testable hypothesis concerning the effect of political polarization. Table 6 investigates the extent to which the effect of political competition is stronger in jurisdictions where the polity is more polarized. As a proxy for political polarization, we use the indicator of ethnic fractionalization from Alesina et al. (2003), which is increasing in the degree of fractionalization.[13] We thus estimate the following:

$$RCT_{i,t} = \alpha PolComp_{i,t} + \beta PolComp_{i,t} \times Ethnic_i + \gamma X_{i,t} + \delta_t + u_{i,t}, \tag{9}$$

where $PolComp_{i,t} \times Ethnic_i$ is the coefficient of interest. As the ethnic fractionalization indicator is not time varying, we drop the country fixed effect from this specification. Our baseline results are strengthened by including the interaction term. Political competition and the interaction term with the proxy for polarization are jointly statistically significant for RCT's in general and, more specifically, for policy learning experiments and public sector accountability experiments. In support of the hypothesis we have theoretically derived, greater political competition increases the likelihood that an experiment is held (above a low threshold of fractionalization) and the marginal effect is greater in more fractionalized societies.[14] Figure 5 provides the marginal effect of an increase in political competition as

[12] We have also constructed analogous tables of results for estimates that employ different measures of political competition. In the Online Appendix, Tables A5 – A8 use Vanhanen's measure of political competition based on ruling party seat margins, making corrections for differences in political institutions. This measure seems to be the best match to the measure of political competition that we employ in our analysis of the Indian states panel data set in the next sub-section. Tables A9 – A11 use the p_parcomp measure of competitiveness of political competition. Results are largely robust to the use of these alternative measure of political competition. Interested readers are referred to the Online Appendix.

[13] Combining racial and linguistic characteristics, the variable measures the probability that two randomly selected people will not share a certain characteristic.

[14] Online Appendix Tables A12 – A14 repeat the exercise with the more narrow accountability categories and with alternative measures of political competition.

Table 6: Experiments and competitiveness of executive recruitment (p_xrcomp)

Panel A	Binary DV's: Fixed effects LPM				
	All RCTs	Policy learning	Electoral learning	Account -ability	Minority inclusion
	(1)	(2)	(3)	(4)	(5)
political competition	-0.101	-0.083	-0.007	-0.059**	-0.004
	(0.06)	(0.06)	(0.01)	(0.03)	(0.01)
pol. comp × ethnic frac.	0.231**	0.185**	0.022	0.185***	0.017
	(0.10)	(0.09)	(0.02)	(0.06)	(0.02)
ln(gdppc)	-0.021	-0.006	-0.003	0.005	0.012
	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)
ln(population)	0.116	0.070	0.029	0.105*	0.026
	(0.10)	(0.09)	(0.03)	(0.06)	(0.02)
within R^2	0.0803	0.0684	0.0086	0.0657	0.0124
joint F-stat p-value	0.0103	0.0259	0.4627	0.0109	0.4893
Panel B		Shares DV's: Fixed effects OLS			
		Policy learning	Electoral learning	Account -ability	Minority inclusion
		(1)	(2)	(3)	(4)
political competition		-0.080	-0.007	-0.065**	-0.000
		(0.05)	(0.01)	(0.03)	(0.00)
pol. comp × ethnic frac.		0.180**	0.022	0.183***	0.003
		(0.08)	(0.02)	(0.06)	(0.00)
ln(gdppc)		-0.007	-0.000	-0.019	-0.000
		(0.03)	(0.00)	(0.02)	(0.00)
ln(population)		0.065	0.028	0.102*	0.010
		(0.09)	(0.03)	(0.06)	(0.01)
within R^2		0.0693	0.0089	0.0503	0.0076
joint F-stat p-value		0.0257	0.4736	0.0125	0.3343
District & year FE's	yes	yes	yes	yes	yes
N	2579	2579	2579	2579	2579
Countries	158	158	158	158	158
N. experiments	283	253	45	94	23

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

a function of the degree of ethnic fractionalization on the probability of running a public sector accountability experiment as estimated in column (4) of panel A.

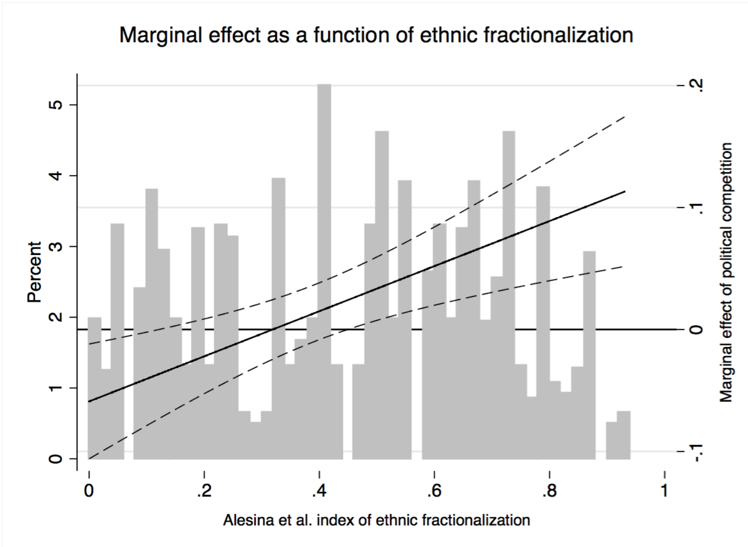

Figure 5: Marginal effect of an increase in political competition as a function of ethnic fractionalization.

5.4 Results with Indian state-level data

In this sub-section, we consider a panel of Indian states from 1995 until 2015. The tables of results that we present for the Indian states panel follows the same structure as the tables for our cross-national investigation. In this case, we use the ruling seats margin enjoyed by the party in control of the state parliaments as our measure of political competition. While the ruling seats margin seems a particularly attractive measure of the probability that the current government will be in power the next period, we note that it is only suitable for an analysis within which the political institutions are the same for all entities, which is why we employ it here, but not in the cross-country analysis.[15]

Table 7 is the analogue of Table 3. Noting that greater ruling seats margins are associated with less political competition, the results from the Indian states panel corroborate those from the cross-country panel. Namely, variation in the level of political competition is not statistically significantly correlated with the likelihood that a state hosts an RCT in general, but it is statistically significantly correlated with the likelihood that a state hosts a public sector accountability experiment in particular. The result has a lower degree of significance than that from the cross-country regressions, which may be due to the lower statistical power of the much reduced sample size.

Table 8 is the analogue of Table 4. The non-linear estimators verify that greater political

[15]We finished the data collection in the spring of 2016. As there are lags between when experiments are began and when the results are written up and published by JPAL, the most recent experiments are not always taken into account (as indicated by the downward kink in the trend line at around 2010 in Figure 5.3). We have re-ran the analysis, dropping observations from the last few years of the panel, and results are very similar.

Table 7: Experiments and political institutions in India

Panel A	Binary DV: Fixed effects LPM regressions				
	All RCTs	Policy learning	Electoral learning	Account -ability	Minority inclusion
	(1)	(2)	(3)	(4)	(5)
Margin of seats	-0.076	-0.088	-0.019	-0.098*	0.013
	(0.12)	(0.14)	(0.03)	(0.06)	(0.04)
ln(gsppc)	0.034	0.002	0.072	-0.035	0.110
	(0.17)	(0.16)	(0.08)	(0.11)	(0.09)
ln(population)	-0.087	-0.147	0.232	-0.109	0.128
	(0.39)	(0.35)	(0.23)	(0.21)	(0.23)
Panel B		**Shares DV: Fixed effects OLS regressions**			
		Policy learning	Electoral learning	Account -ability	Minority inclusion
		(1)	(2)	(3)	(4)
Margin of seats		-0.083	-0.012	-0.052*	-0.000
		(0.14)	(0.02)	(0.03)	(0.05)
ln(gsppc)		-0.024	0.037	-0.027	0.130
		(0.15)	(0.05)	(0.05)	(0.11)
ln(population)		-0.215	0.157	-0.116	0.192
		(0.33)	(0.16)	(0.14)	(0.27)
District & year FE's	yes	yes	yes	yes	yes
N	553	553	553	553	553
Districts	30	30	30	30	30
N. experiments	95	87	12	27	17

Notes: Robust standard errors clustered by Indian state are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

competition increases the likelihood of hosting a governmental accountability experiment in Indian states. The also indicate that political competition affects the likelihood of hosting a policy learning experiment and a RCT more generally.

Table 9 investigates the robustness of the result from model B3 of Table 7 by adding additional time-varying controls to the fixed-effects panel regression. The structure of the table follows that of the cross-country investigation in Table 5. We control for whether it was an election year, some proxies for state capacity, a proxy of political awareness, and the urbanized population. Coefficient estimates on the ruling margin of seats is very stable across these specifications.[16]

[16]Tables A15 – A17 in the Online Appendix confirms that the results are robust to using the Herfindahl-Hirschman Index of seat concentration as a measure of political competition. Table A18 estimates random

Table 8: Experiments and political institutions in India

	Binary dependent variables							
	Fixed effects logit regressions				Random effects logit regressions			
	All RCTs	Policy learning	Electoral learning	Account -ability	All RCTs	Policy learning	Electoral learning	Account -ability
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
margin_seats	-2.102**	-2.082**	0.685	-3.788**	-1.824*	-1.910**	2.967	-3.117*
	(1.02)	(1.01)	(8.96)	(1.65)	(0.97)	(0.96)	(1.80)	(1.65)
ln(gsppc)	-0.775	-0.675	-6.792*	-0.388	0.203	0.343*	-0.544	0.156
	(0.69)	(0.71)	(3.93)	(1.16)	(0.21)	(0.21)	(0.40)	(0.36)
ln(population)	8.969	9.145	47.288	7.352	1.617***	1.511***	1.808***	3.073**
	(5.67)	(5.88)	(34.99)	(10.30)	(0.39)	(0.36)	(0.57)	(1.32)
District FE's	yes	yes	yes	yes	no	no	no	no
N	266	266	40	137	553	553	553	553
Districts	14	14	2	7	30	30	30	30
N. experiments	95	87	12	27	95	87	12	27

Notes: *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table 9: Political accountability experiments and political competition in India

	Share accountability experiments DV: Fixed effects OLS regressions						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Margin seats	-0.047*	-0.052*	-0.051*	-0.056*	-0.047*	-0.046*	-0.053*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
ln(gsppc)		-0.027	-0.017	-0.025	-0.019	-0.009	-0.015
		(0.05)	(0.06)	(0.06)	(0.07)	(0.06)	(0.08)
ln(population)		-0.116	-0.106	-0.174	-0.141	-0.085	-0.267
		(0.14)	(0.25)	(0.32)	(0.21)	(0.23)	(0.40)
state election		0.008					0.000
		(0.01)					(0.00)
tax revenues			0.002				0.003
			(0.01)				(0.01)
fiscal deficit				-0.000			-0.000
				(0.00)			(0.00)
literacy rate					0.379		0.496
					(0.51)		(0.57)
urbanization rate						0.001	0.002
						(0.00)	(0.00)
District & year FE's	yes	yes	yes	yes	yes	yes	yes
N	591	553	495	474	435	435	399
Districs	30	30	30	30	30	30	30
N. experiments	27	27	27	27	26	26	26

Notes: Robust standard errors clustered by Indian state are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

effects Poisson models using count dependent variables.

6 Conclusion

This paper deals with the political economy of RCTs for evaluation of international development policies by considering the political environment in which RCTs are more likely to occur. Focusing on a "tying the hands" mechanism, we theorize that the higher the probability of loosing power (i.e., the more competitive the political environment), the more likely is the incumbent government to devote some resources to learning how to constrain the future government. In our model, running an experiment increases the quality design of a reform and allows the incumbent to learn more about the concrete effects of a reform. To the extent that it is government actors that are motivated by political competition to run experiments, we should observe that political competition only affects the likelihood of experiments on which the government is an explicit partner. We also demonstrate that the effect of political competition on the likelihood to run an experiment is magnified by the degree of political polarization in the jurisdiction, as the utility loss of loosing power and of being in the opposition group is higher in a polarized society.

We then mobilize an original dataset on RCTs concerning international development that have been conducted all other the world and within Indian states, which constitute an important share of experiments in the world. Our three main results support the theoretical hypotheses derived from our formal analysis. First, we show that the probability of observing a public sector accountability experiment increases substantially with our proxy for political competition, though there is no unconditional effect on the probability of observing an experiment in general. Second, political competition affects the likelihood that an RCT with government involvement is held, but does not affect the likelihood that an RCT is run independent of the domestic government. Third, we observe a much stronger effect of political competition in ethnically polarized countries, consistent with the "tying the hands" mechanism that we highlight in our model. When allowing for heterogenous affects, we show that political competition affects the likelihood of hosting an experiment, in general, not only for public sector accountability experiments.

Our results suggest that experiments concerning public sector accountability occur in a very specific political environment and that the result obtained from those experiments may not be universal given that environments characterized by high political competition may be very different from other environments which could effect the results. Of course, the political site selection bias that we have identified in this paper is only one of many potential selection biases that may bring into question the external validity of such program evaluations of policies for international development. Even when aggregating a large collection of experiments, the fact that some categories of RCTs occur in a specific environment should make researchers careful in drawing general conclusions and policy recommendations from their results. For future research in this domain, we see two promising avenues. First, research should continue to investigate the specific environment in which experiments occur, political competition being only one dimension among potentially many others. Second research should question how the specific environment we have documented precisely translates into the external validity of experiments when interpreting the results. We leave these questions for future research.

References

Acemoglu, D., 2010. Theory, general equilibrium, and political economy in development economics. Journal of Economic Perspectives 24 (3), 17–32.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., Wacziarg, R., 2003. Fractionalization. Journal of Economic Growth 8 (3), 155 – 194.

Allcott, H., 2015. Site selection bias in program evaluation. The Quarterly Journal of Economics 130 (3), 1117–1165.

Allcott, H., Mullainathan, S., 2012. External validity and partner selection bias. Tech. rep.

Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., Walton, M., 2017. From proof of concept to scalable policies: Challenges and solutions, with an application. Journal of Economic Perspectives 31 (4), 73–102.

Banerjee, A., Duflo, E., Imbert, C., Mathew, S., Pande, R., 2016. E-governance, accountability, and leakage in public programs: Experimental evidence from a financial management reform in India. Tech. rep., National Bureau of Economic Research.

Banerjee, A. V., Duflo, E., 2014. Do firms want to borrow more? Testing credit constraints using a directed lending program. Review of Economic Studies 81 (2), 572–607.

Besley, T., Persson, T., Reynal-Querol, M., 2016. Resilient leaders and institutional reform: Theory and evidence. Economica 83 (332), 584–623.

Blair, G., Iyengar, R. K., Shapiro, J. N., 2013. Where policy experiments are conducted in economics and political science: The missing autocracies. Tech. rep., Working Paper, Princeton University (May).

Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., Sandefur, J., 2013. Interventions & institutions experimental evidence on scaling up education reforms in Kenya. CSAE Working Paper WPS/2013-04.

Briceño, B., Cuesta, L., Attanasio, O., 2011. Behind the scenes: Experience managing and conducting large impact evaluations in Colombia. Journal of Development Effectiveness 3 (4), 470–501.

Chassang, S., Snowberg, E., et al., 2012. Selective trials: A principal-agent approach to randomized controlled experiments. The American Economic Review 102 (4), 1279–1309.

De Figueiredo, R. J., 2002. Electoral competition, political uncertainty, and policy insulation. American Political Science Review 96 (2), 321–333.

Deaton, A., 2010. Instruments, randomization, and learning about development. Journal of Economic Literature 48 (2), 424–455.

Deaton, A., Cartwright, N., 2016. Understanding and misunderstanding randomized controlled trials. Tech. rep., National Bureau of Economic Research.

Dehejia, R., Pop-Eleches, C., Samii, C., 2015. From local to global: External validity in a fertility natural experiment. Tech. rep., National Bureau of Economic Research.

Duflo, E., 2004. Scaling up and evaluation. In: Fran c. B., Pleskovic, B. (Eds.), Accelerating Development: Annual World Bank Conference on Development Economics. The World Bank, Washington DC, pp. 341–369.

Dunning, T., 2016. Transparency, replication, and cumulative learning: What experiments alone cannot achieve. Annual Review of Political Science 19, 541–563.

Dunning, T., Grossman, G., Humphreys, M., Hyde, S., McIntosh, C., 2017. Information and Accountability: A New Method for Cumulative Learning. Cambridge University Press.

Faulkner, W. N., 2014. A critical analysis of a randomized controlled trial evaluation in Mexico: Norm, mistake or exemplar? Evaluation 20 (2), 230–243.

Geddes, B., 1994. Politician's Dilemma: Building State Capacity in Latin America. University of California Press.

Green, D. P., Gerber, A. S., 2002. Reclaiming the Experimental Tradition in Political Science. WW Norton and Company.

Green, D. P., Gerber, A. S., 2015. Get Out the Vote: How to Increase Voter Turnout. Brookings Institution Press, Washington DC.

Grose, C. R., 2014. Field experimental work on political institutions. Annual Review of Political Science 17, 355–370.

Grzymala-Busse, A., 2007. Rebuilding Leviathan: Party Competition and State Exploitation in Post-Communist Democracies. Cambridge University Press.

Gueron, J. M., 2017. The politics and practice of social experiments: Seeds of a revolution. In: Banerjee, A., Duflo, E. (Eds.), Handbook of Economic Field Experiments. Vol. 1. Elsevier, pp. 27–69.

Gueron, J. M., Rolston, H., 2013. Fighting for Reliable Evidence. Russell Sage Foundation.

Humphreys, M., Weinstein, J. M., 2009. Field experiments and the political economy of development. Annual Review of Political Science 12, 367–378.

Imbens, G. W., 2010. Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). Journal of Economic Literature 48 (2), 399–423.

Lawson, M. L., 2012. Does Foreign Aid Work?: Efforts to Evaluate US Foreign Assistance. Congressional Research Service.

Levy, S., 2006. Progress against poverty. Washington DC: Brookings Institute.

Lipsky, D. B., Seeber, R. L., Avgar, A. C., Scanza, R. M., 2007. Managing the politics of evaluation: Lessons from the evaluation of ADR programs. Proceedings of the Fifty-Ninth Annual Meeting of the Labor and Employment Relations Association, 116–129.

Loewen, P. J., Rubenson, D., Wantchekon, L., 2010. Help me help you: Conducting field experiments with political elites. The ANNALS of the American Academy of Political and Social Science 628 (1), 165–175.

Marshall, M. G., Jaggers, K., Gurr, T. R., 2016. Polity IV Project, Political Regime Characteristics and Transitions, 1800 – 2014.

Miguel, E., Kremer, M., 2004. Worms: Identifying impacts on education and health in the presence of treatment externalities. Econometrica 72 (1), 159–217.

Moe, T. M., 1989. The politics of bureaucratic structure. In: Chubb, J., Peterson, P. (Eds.), Can the Government Govern? The Brookings Institute, Washington DC, pp. 267 – 329.

Muralidharan, K., Niehaus, P., Sukhtankar, S., 2016. Building state capacity: Evidence from biometric smartcards in India. The American Economic Review 106 (10), 2895–2929.

Pritchett, L., 2002. It pays to be ignorant: A simple political economy of rigorous program evaluation. The Journal of Policy Reform 5 (4), 251–269.

Vanhanen, T., 2016. Measures of Democracy, 1810 – 2012.

Vivalt, E., 2016. How much can we generalize from impact evaluations? Unpublished Manuscript: Stanford University.

Wantchekon, L., 2003. Clientelism and voting behavior: Evidence from a field experiment in Benin. World politics 55 (3), 399–422.

Wantchekon, L., Guardado R, J., 2011. Methodology update: Randomised controlled trials, structural models and the study of politics. Journal of African Economies 20 (4), 653–672.

Weiss, C. H., 1993. Where politics and evaluation research meet. Evaluation Practice 14 (1), 93–106.

A Appendix of Proofs

Proof of Proposition 1

First, note that when $\zeta = 0$, we have $V_I\left(R = 1; E = 1; C_O = C_O^h; \zeta = 0\right) >$ $V_I\left(R = 1; E = 1; C_O = C_O^l; \zeta = 0\right) > V_I\left(R = 0; E = 1; \zeta = 0\right)$. The value of reforming is higher when $C_O = C_O^h$ and the value of reforming is always higher than the value of not reforming. When $\zeta = 1$, we have $V_I\left(R = 1; E = 1; C_O = C_O^h; \zeta = 1\right) =$ $V_I\left(R = 1; E = 1; C_O = C_O^l; \zeta = 1\right) < V_I\left(R = 0; E = 1; \zeta = 1\right)$. It is never profitable to implement a policy change when $\zeta = 1$. As a result, $V_I\left(R = 1; E = 1; C_O = C_O^h; \zeta = 0\right)$ and $V_I\left(R = 1; E = 1; C_O = C_O^l; \zeta = 0\right)$ should intersect $V_I\left(R = 0; E = 1; \zeta = 0\right)$ uniquely respectively for $\zeta = \overline{\zeta}$ and $\zeta = \underline{\zeta}$ given they are linear functions with respect to ζ. Since $V_I\left(R = 0; E = 1; \zeta = 0\right)$ and $V_I\left(R = 1; E = 1; C_O = C_O^i\right)$ are strictly increasing in ζ and $V_I\left(R = 1; E = 1; C_O = C_O^h\right) > V_I\left(R = 1; E = 1; C_O = C_O^l\right)$ for any $\zeta < 1$ then we have $\overline{\zeta} > \underline{\zeta}$. The threshold values $\underline{\zeta} = C_O^l/(C_O^l + C_I)$ and $\overline{\zeta} = C_O^h/(C_O^h + C_I)$ are respectively the solutions of $V_I\left(R = 1; E = 1; C_O = C_O^l\right) = V_I\left(R = 0; E = 1\right)$ and $V_I\left(R = 1; E = 1; C_O = C_O^l\right) = V_I\left(R = 0; E = 1\right)$. We can verify that $\underline{\zeta} = C_O^l/(C_O^l + C_I) < \overline{\zeta} = C_O^h/(C_O^h + C_I)$ given $C_O^l < C_O^h$. Q.E.D.

Proof of Proposition 2

First when $\zeta = 0$ we can see that $V_I\left(R = 0; E = 0; \zeta = 0\right) < V_I\left(R = 1; E = 0; \zeta = 0\right)$ and when $\zeta = 1$, $V_I\left(R = 0; E = 0; \zeta = 1\right) > V_I\left(R = 1; E = 0; \zeta = 1\right)$. Given both value functions $V_I\left(R = 0; E = 0\right)$ and $V_I\left(R = 1; E = 0\right)$ are strictly linear in ζ they should intersect for a unique value of ζ. The threshold values $\zeta_c = E(C_O)/(E(C_O) + C_I)$ is the solutions of $V_I\left(R = 1; E = 0\right) = V_I\left(R = 0; E = 0\right)$. We can verify that $\zeta_c \in \left[\underline{\zeta}, \overline{\zeta}\right]$ since $\underline{\zeta} = C_O^l/(C_O^l + C_I)$ and $\overline{\zeta} = C_O^h/(C_O^h + C_I)$ and given that $C_O^l < E(C_O) < C_O^h$. Q.E.D.

Proof of Proposition 3

First we can see that when $\zeta = 0$ we have $V_I\left(E = 1; \zeta = 0\right) > V_I\left(R = 0; E = 0; \zeta = 0\right)$ if and only if $\beta P_E E(C_O) > c^e$ that is when the cost of running an experiment is not to high. Second, we can see that when $\zeta > \overline{\zeta}$ we always have $V_I\left(E = 1, \zeta > \overline{\zeta}\right) < V_I\left(R = 0; E = 0\right)$ for $c^e > 0$. Since $V_I\left(E = 1\right)$ is an increasing function of ζ for $\zeta > \overline{\zeta}$, $\underline{\zeta} < \zeta < \overline{\zeta}$ and $\zeta < \underline{\zeta}$ and is continuous in ζ, it should intersect $V_I\left(R = 0; E = 0\right)$ once in the $\left[0, \overline{\zeta}\right]$ space. The threshold value $\zeta_E = (E(C_O) - (c^e/\beta P_E))/(E(C_O) + C_I)$ is the solution of $V_I\left(E = 1, \zeta < \underline{\zeta}\right) = V_I\left(R = 0; E = 0\right)$ and $\zeta_E = (C_O^h - (c^e/\beta P_E))/(C_O^h + C_I)$ is the solution of $V_I\left(E = 1, \overline{\zeta} > \zeta > \underline{\zeta}\right) = V_I\left(R = 0; E = 0\right)$. We ca see $\zeta_E = (C_O^h - (c^e/\beta P_E))/(C_O^h + C_I) < \overline{\zeta}$

when $c^e > 0$. Q.E.D.

Proof of Proposition 4

When $\zeta = 0$ and $c^e = 0$, we can see that $V_I\left(E = 1, \zeta < \underline{\zeta}\right) > V_I\left(R = 1; E = 0\right) > V_I\left(R = 0; E = 0\right)$. When $\zeta \geq \overline{\zeta}$ and $c^e = 0$ we can see that $V_I(E = 1, \zeta \geq \overline{\zeta}) = V_I\left(R = 0; E = 0\right) > V_I\left(R = 1; E = 0\right)$. Given that $V_I\left(R = 1; E = 0\right) = V_I\left(R = 0; E = 0\right)$ for the value ζ_c which is lower than $\overline{\zeta}$ (see proposition 2) and that $V_I(E = 1, \zeta \geq \overline{\zeta}) = V_I\left(R = 0; E = 0\right)$ when $\zeta \geq \overline{\zeta}$ then $V_I\left(E = 1, \zeta < \overline{\zeta}\right) > V_I\left(R = 1; E = 0\right)$ for all $\zeta < \overline{\zeta}$ (see the right panel of figure 2 for the case where $c^e = 0$). Q.E.D.

B Supplementary Information about the Data

We selected all the J-PAL experiments in political economy and governance, health and education conducted between 1995 and 2015 and included in the online repository as primary areas of interest in international development. In order to match the theoretical model, we recoded the RCT experiments according to parameters reflecting the benefits and costs for incumbents and future opponents in six categories, as follows:

Electoral learning consists in experiments whereby bureaucrats and politicians test the most efficient ways to appeal to voters. The treatments in this category include alternative communication channels for electoral messages (mail, canvassing, online ads) for various groups of voters, or message content (framing, programmatic versus clientelistic appeals, etc.). The major characteristic of RCTs in this category is that they help political campaigns and/or researchers learn how to better appeal to their base, but do not entail indirect costs for incumbent politicians, opponents, or bureaucrats other than the administrative cost of RCT implementation (High benefit, low cost).

Policy learning contains the experiments that entail precise tests of the impact of a policy on an technical outcome of interest. The examples of this category are heterogeneous and include RCTs as diverse as testing deworming treatments on school attendance all the way to the role of internal auditors in corporate performance and governance. While these RCTs offer clear benefits to decision-makers and researchers in terms of increasing the probability of policy adoption (Pe), the costs for the incumbent and opponents depend on the distance between the RCT results and their ideal points that are policy and context specific (High benefit, costs varying according to policy distance from ideal point).

Minority inclusion experiments include the RCTs of our dataset that attempt to find the most effective ways to include gender, racial, ethnic, linguistic minorities, or socio-economic groups traditionally disempowered in formal or informal decision-making institutions, or that assess the impact of specific policy tools of minority representation on the general performance of the members of minority groups. Examples include election quotas for women and their impact on representation, as well as the impact of minority leaders on attitudes of members of the minority group. Additionally, this category also contains experiments that assess the impact of diversity on outcomes such as social cooperation and public good production. Like the previous category, this family of experiments entails learning benefits for researchers and policy makers, while the indirect costs are likely to vary according to the ideological distance between results and policy ideal points.

Political accountability experiments comprise RCTs that directly test policy tools that could increase the responsiveness of public officials to citizens, voters and public service users. This family of experiments aims at reducing political corruption, increase the transparency of public accounts, and assess the effectiveness of various auditing mechanisms. The characteristic of this type of RCTs is that the indirect costs for both incumbent politicians and opponents are high since they propose the best mechanisms to place constraints on political discretion. The direct benefit for the politician could be also high given the fact such a constraint on the executive will also benefit the opposition group the current incumbent could belonging in the futur. (High benefit, High costs).

Bureaucratic accountability is a related category of RCTs that test the best ways to reduce bureaucratic rather than political discretion. Examples of treatments include monitoring tools and professional incentives for public sector teachers and healthcare workers attendance and performance, tax collector probity vis--vis taxpayers, bureaucratic transparency, the identification and elimination of ghost workers from public sector payrolls, as well as simplification and streamlining in transfers and public service delivery. Such experiments bear elevated costs for corrupt bureaucrats and politicians who have the incentive to staff the bureaucracy with political partisans as a form of patronage politics, but bring benefits in terms of informing the most effective policies that could tie the hands of future political opponents coming to power. (High benefit, high costs)

Decentralization, our last category of RCTs, refers to the family of experiments that test the effect of the total or partial devolution of certain public services to non-state entities such as NGOs, village committees, community volunteers, parent teaching associations, etc. This family of experiments also includes evaluations of the impact of shared governance arrangements of state and non-state groups on policy outcomes. We hypothesize that the costs for incumbent and opposition politicians are similar to those of the accountability experiments category (High costs for the incumbent and opponent).

C Appendix of Tables

C.1 Data summaries

Table A1: Summary statistics for the cross-country panel

variable	N	mean	std. dev.	min	max
randomized controlled trial (binary)	3205	0.089	0.285	0	1
policy learning experiment (binary)	3205	0.079	0.271	0	1
policy learning experiment (share)	3205	0.077	0.264	0	1
electoral learning experiment (binary)	3205	0.014	0.119	0	1
electoral learning experiment (share)	3205	0.008	0.079	0	1
government accountability experiment (binary)	3205	0.030	0.170	0	1
government accountability experiment (share)	3205	0.027	0.189	0	1
minority inclusion experiment (binary)	3205	0.007	0.084	0	1
minority inclusion experiment (share)	3205	0.003	0.044	0	1
p_xrcomp	2607	2.023	1.077	0	3
e_van_comp	2780	39.444	22.459	0	70
p_parcomp	2607	3.335	1.353	0	5
ethnic fractionalization	3143	0.440	0.257	0	0.930
gross domestic production per capita (logged)	3205	8.044	1.645	4.151	12.109
population (logged)	3205	15.410	2.222	9.134	21.043
legislative election (binary)	2899	0.220	0.414	0	1
tax revenue as percentage of GDP	1818	16.962	8.320	0.020	65.903
public education as percentage of GDP	1738	4.615	1.910	0.689	15.615
urbanization rate	3188	53.888	23.588	7.412	100

Notes: Summary statistics are calculated over the panel of countries for which we have fh_ipolity2 data.

Table A2: Summary statistics for Indian data

variable	N	mean	std. dev.	min	max
randomized controlled trial (binary)	692	0.140	0.347	0	1
policy learning experiment (binary)	692	0.129	0.335	0	1
policy learning experiment (share)	692	0.121	0.322	0	1
electoral learning experiment (binary)	692	0.017	0.131	0	1
electoral learning experiment (share)	692	0.012	0.101	0	1
government accountability experiment (binary)	692	0.039	0.194	0	1
government accountability experiment (share)	692	0.020	0.108	0	1
minority inclusion experiment (binary)	692	0.027	0.164	0	1
minority inclusion experiment (share)	692	0.026	0.165	0	1
margin of majority, by seats	561	0.297	0.222	0	1
HH index of seat concentration	561	0.772	0.218	0	1.080
gross state production per capita (logged)	620	10.405	0.814	8.134	12.459
state population (logged)	688	15.906	2.112	10.924	19.185
state election (binary)	690	0.172	0.378	0	1
tax revenue raised as percentage of GSP	512	5.723	2.333	0	11.76
fiscal deficit as percentage of GSP	481	-4.892	12.016	-256.340	7.31
literacy rate	527	0.620	0.105	0 .37	0.84
urbanization rate	560	31.356	17.010	8.852	93.176

C.2 Further results: cross-national data

Table A3: Experiments and competitiveness of executive recruitment across countries, without the advanced industrialized democracies

Panel A	Binary dependent variables Fixed effects LPM regressions				
	All RCTs	Policy learning	Electoral learning	Account-ability	Minority inclusion
	(1)	(2)	(3)	(4)	(5)
p_xrcomp	0.033	0.024	0.005	0.046**	0.005
	(0.02)	(0.02)	(0.00)	(0.02)	(0.00)
ln(gdppc)	-0.024	-0.008	-0.006	-0.016	-0.001
	(0.03)	(0.03)	(0.00)	(0.02)	(0.00)
ln(population)	0.105	0.062	0.021	0.060	0.007
	(0.11)	(0.10)	(0.03)	(0.06)	(0.00)
Panel B		**Shares dependent variables Fixed effects OLS regressions**			
		Policy learning	Electoral learning	Account-ability	Minority inclusion
		(1)	(2)	(3)	(4)
p_xrcomp		0.025	0.005	0.040**	0.001
		(0.02)	(0.00)	(0.02)	(0.00)
ln(gdppc)		-0.006	-0.004	-0.030	-0.002
		(0.03)	(0.00)	(0.02)	(0.00)
ln(population)		0.061	0.022	0.074	0.007
		(0.09)	(0.03)	(0.06)	(0.01)
District & year FE's	yes	yes	yes	yes	yes
N	2199	2199	2199	2199	2199
Countries	136	136	136	136	136

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A4: Experiments and competitiveness of executive recruitment across countries

Panel A	Binary dependent variables Fixed effects LPM regressions						
	All RCTs	Policy learning	Electoral learning	Political account	Bureau account	Decentral -ization	Minority inclusion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
p_xrcomp	0.033	0.024	0.006	0.012	0.030**	0.027	0.006
	(0.02)	(0.02)	(0.00)	(0.01)	(0.01)	(0.02)	(0.00)
ln(gdppc)	-0.021	-0.006	-0.002	-0.000	0.021	-0.018*	0.012
	(0.03)	(0.03)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
ln(population)	0.120	0.073	0.029	0.024	0.063	0.053	0.027
	(0.10)	(0.09)	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)
Panel B		Shares dependent variables Fixed effects OLS regressions					
		Policy learning	Electoral learning	Political account	Bureau account	Decentral -ization	Minority inclusion
		(1)	(2)	(3)	(4)	(5)	(6)
p_xrcomp		0.025	0.006	0.010	0.017**	0.014	0.001
		(0.02)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
ln(gdppc)		-0.007	-0.000	-0.006	0.005	-0.017**	0.000
		(0.03)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
ln(population)		0.068	0.029	0.026	0.034	0.048	0.010
		(0.09)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)
Country & year FE's	yes	yes	yes	yes	yes	yes	yes
N	2607	2607	2607	2607	2607	2607	2607
Countries	160	160	160	160	160	160	160

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A5: Experiments and political competition across countries (Vanhanen 2016 measure)

Panel A	Binary dependent variables Fixed effects LPM regressions				
	All RCTs	Policy learning	Electoral learning	Account -ability	Minority inclusion
	(1)	(2)	(3)	(4)	(5)
e_van_comp	0.002**	0.001*	-0.000	0.001*	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ln(gdppc)	-0.014	-0.002	0.000	0.008	0.010
	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)
ln(population)	0.204*	0.119	0.051	0.175**	0.029
	(0.12)	(0.10)	(0.04)	(0.09)	(0.02)
Panel B		Share dependent variables Fixed effects OLS regressions			
		Policy learning	Electoral learning	Account -ability	Minority inclusion
		(1)	(2)	(3)	(4)
e_van_comp		0.001*	-0.000	0.002**	-0.000
		(0.00)	(0.00)	(0.00)	(0.00)
ln(gdppc)		-0.002	0.002	-0.014	0.000
		(0.03)	(0.00)	(0.02)	(0.00)
ln(population)		0.113	0.051	0.184**	0.014
		(0.10)	(0.04)	(0.09)	(0.01)
Country & year FE's	yes	yes	yes	yes	yes
N	2780	2780	2780	2780	2780
Countries	164	164	164	164	164
N. experiments	286	255	46	95	23

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A6: Experiments and political competition across countries (Vanhanen 2016 measure)

Panel A	Binary dependent variables Fixed effects LPM regressions						
	All RCTs	Policy learning	Electoral learning	Political account	Bureau account	Decentral -ization	Minority inclusion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
e_van_comp	0.002**	0.001*	-0.000	0.000	0.001**	0.001	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ln(gdppc)	-0.014	-0.002	0.000	0.002	0.019	-0.016	0.010
	(0.03)	(0.03)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
ln(population)	0.204*	0.119	0.051	0.057	0.079	0.081*	0.029
	(0.12)	(0.10)	(0.04)	(0.05)	(0.05)	(0.05)	(0.02)
Panel B		Shares dependent variables Fixed effects OLS regressions					
		Policy learning	Electoral learning	Political account	Bureau account	Decentral -ization	Minority inclusion
		(1)	(2)	(3)	(4)	(5)	(6)
e_van_comp		0.001*	-0.000	0.000	0.001*	0.001	-0.000
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ln(gdppc)		-0.002	0.002	-0.003	0.005	-0.016**	0.000
		(0.03)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
ln(population)		0.113	0.051	0.059	0.050	0.075*	0.014
		(0.10)	(0.04)	(0.05)	(0.04)	(0.04)	(0.01)
Country & year FE's	yes	yes	yes	yes	yes	yes	yes
N	2780	2780	2780	2780	2780	2780	2780
Countries	164	164	164	164	164	164	164
N. experiments	286	255	46	45	46	46	23

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A7: Political accountability experiments and political competition across countries (Vanhanen 2016 measure)

	Share accountability experiments Fixed effects OLS regressions					
	(1)	(2)	(3)	(4)	(5)	(6)
e_van_comp	0.002**	0.002**	0.004***	0.002	0.002**	0.004**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ln(gdppc)		-0.021	-0.021	-0.034	-0.013	-0.058
		(0.02)	(0.03)	(0.03)	(0.02)	(0.04)
ln(population)		0.186**	0.311**	0.188	0.175*	0.129
		(0.09)	(0.15)	(0.12)	(0.10)	(0.21)
parliamentary election		-0.004				-0.001
		(0.00)				(0.01)
tax revenues			0.000			-0.003
			(0.00)			(0.00)
education expenditures				0.014		0.024
				(0.01)		(0.02)
urbanization					0.002	0.004
					(0.00)	(0.01)
Country & year FE's	yes	yes	yes	yes	yes	yes
N	2780	2729	1657	1616	2780	1085
Countries	192	172	156	176	188	131

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A8: Experiments and political competition across countries (Vanhanen 2016 measure)

	Binary dependent variables							
	Fixed effects logit regressions				Random effects logit regressions			
	All RCTs	Policy learning	Electoral learning	Account -ability	All RCTs	Policy learning	Electoral learning	Account -ability
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
e_van_comp	0.020*	0.019*	0.003	0.044***	0.024***	0.025**	0.004	0.038***
	(0.01)	(0.01)	(0.04)	(0.02)	(0.01)	(0.01)	(0.02)	(0.01)
ln(gdppc)	2.185***	2.522***	0.435	5.109***	1.505***	1.548***	0.092	1.149***
	(0.84)	(0.86)	(1.55)	(1.50)	(0.29)	(0.32)	(0.34)	(0.37)
ln(population)	3.629	2.220	13.944**	2.311	3.924***	3.610***	1.977***	4.329***
	(3.08)	(3.24)	(6.88)	(4.67)	(0.61)	(0.64)	(0.66)	(0.65)
Country FE's	yes	yes	yes	yes	no	no	no	no
Year FE's	yes	yes	yes	yes	yes	yes	yes	yes
N	731	680	170	340	2780	2780	2780	2780
Countries	43	40	10	20	164	164	164	164
N. experiments	269	238	46	95	269	238	46	95

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A9: Experiments and competitiveness of political participation across countries

Panel A	Binary dependent variables Fixed effects LPM regressions				
	All RCTs	Policy learning	Electoral learning	Account -ability	Minority inclusion
	(1)	(2)	(3)	(4)	(5)
p_parcomp	0.046***	0.034**	0.000	0.028**	0.002
	(0.02)	(0.01)	(0.00)	(0.01)	(0.00)
ln(gdppc)	-0.028	-0.011	-0.002	0.001	0.012
	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)
ln(population)	0.104	0.061	0.029	0.102	0.027
	(0.10)	(0.09)	(0.03)	(0.06)	(0.02)
Panel B		**Share dependent variables Fixed effects OLS regressions**			
		Policy learning	Electoral learning	Account -ability	Minority inclusion
		(1)	(2)	(3)	(4)
p_parcomp		0.032**	0.000	0.036**	0.001
		(0.01)	(0.00)	(0.02)	(0.00)
ln(gdppc)		-0.012	-0.000	-0.023	-0.000
		(0.03)	(0.00)	(0.02)	(0.00)
ln(population)		0.057	0.029	0.095	0.009
		(0.09)	(0.03)	(0.06)	(0.01)
Country & year FE's	yes	yes	yes	yes	yes
N	2607	2607	2607	2607	2607
Countries	160	160	160	160	160

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A10: Experiments and competitiveness of political participation across countries

Panel A	Binary dependent variables Fixed effects LPM regressions						
	All RCTs	Policy learning	Electoral learning	Political account	Bureau account	Decentral -ization	Minority inclusion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
p_parcomp	0.046***	0.034**	0.000	0.012	0.015*	0.021*	0.002
	(0.02)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
ln(gdppc)	-0.028	-0.011	-0.002	-0.002	0.019	-0.021*	0.012
	(0.03)	(0.03)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
ln(population)	0.104	0.061	0.029	0.020	0.059	0.046	0.027
	(0.10)	(0.09)	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)
Panel B		Shares dependent variables Fixed effects OLS regressions					
		Policy learning	Electoral learning	Political account	Bureau account	Decentral -ization	Minority inclusion
		(1)	(2)	(3)	(4)	(5)	(6)
p_parcomp		0.032**	0.000	0.010	0.011*	0.015	0.001
		(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
ln(gdppc)		-0.012	-0.000	-0.007	0.004	-0.020**	-0.000
		(0.03)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
ln(population)		0.057	0.029	0.022	0.030	0.043	0.009
		(0.09)	(0.03)	(0.03)	(0.02)	(0.03)	(0.01)
Country & year FE's	yes	yes	yes	yes	yes	yes	yes
N	2607	2607	2607	2607	2607	2607	2607
Countries	160	160	160	160	160	160	160

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A11: Political accountability experiments and competitiveness of political participation across countries

	Share accountability experiments Fixed effects OLS regressions					
	(1)	(2)	(3)	(4)	(5)	(6)
p_parcomp	0.026**	0.036**	0.070**	0.018	0.036**	0.048
	(0.01)	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)
ln(gdppc)		-0.023	-0.040	-0.037	-0.021	-0.059
		(0.02)	(0.03)	(0.03)	(0.02)	(0.05)
ln(population)		0.095	0.242**	0.160*	0.081	0.065
		(0.06)	(0.10)	(0.10)	(0.06)	(0.16)
parliamentary election		-0.007*				-0.003
		(0.00)				(0.01)
tax revenues			0.000			-0.003
			(0.00)			(0.00)
education expenditures				0.020*		0.034*
				(0.01)		(0.02)
urbanization					0.004	0.007
					(0.00)	(0.01)
Country & year FE's	yes	yes	yes	yes	yes	yes
N	3121	2607	1551	1533	2607	1037
Countries	163	160	136	148	160	120

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A12: Experiments and competitiveness of executive recruitment across countries

Panel A	Binary dependent variables Fixed effects LPM regressions						
	All RCTs	Policy learning	Electoral learning	Political account	Bureau account	Decentral -ization	Minority inclusion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
p_xrcomp	-0.101	-0.083	-0.007	-0.027	-0.027*	-0.047*	-0.004
	(0.06)	(0.06)	(0.01)	(0.02)	(0.02)	(0.03)	(0.01)
p_xrcomp $\times$ ethnic frac.	0.231**	0.185**	0.022	0.067*	0.098**	0.127**	0.017
	(0.10)	(0.09)	(0.02)	(0.03)	(0.04)	(0.06)	(0.02)
lngdppc	-0.021	-0.006	-0.003	-0.000	0.021	-0.019*	0.012
	(0.03)	(0.03)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
lnpop	0.116	0.070	0.029	0.022	0.061	0.049	0.026
	(0.10)	(0.09)	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)
within R^2	0.0803	0.0684	0.0086	0.0239	0.0386	0.0515	0.0124
joint F-stat p-value	0.0103	0.0259	0.4627	0.1588	0.0293	0.1016	0.4893

Panel B		Shares dependent variables Fixed effects OLS regressions					
		Policy learning	Electoral learning	Political account	Bureau account	Decentral -ization	Minority inclusion
		(1)	(2)	(3)	(4)	(5)	(6)
p_xrcomp		-0.080	-0.007	-0.021	-0.016*	-0.028*	-0.000
		(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
p_xrcomp $\times$ ethnic frac.		0.180**	0.022	0.054*	0.056**	0.073**	0.003
		(0.08)	(0.02)	(0.03)	(0.02)	(0.03)	(0.00)
ln(gdppc)		-0.007	-0.000	-0.006	0.005	-0.018**	-0.000
		(0.03)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
ln(population)		0.065	0.028	0.024	0.032	0.045	0.010
		(0.09)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)
within R^2		0.0693	0.0089	0.0216	0.0344	0.0353	0.0076
joint F-stat p-value		0.0257	0.4736	0.1563	0.0601	0.0726	0.3343
Country & year FE's	yes	yes	yes	yes	yes	yes	yes
N	2579	2579	2579	2579	2579	2579	2579
Countries	158	158	158	158	158	158	158

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A13: Experiments and political competition across countries (Vanhanen 2016 measure)

Panel A	Binary dependent variables Fixed effects LPM regressions				
	All RCTs	Policy learning	Electoral learning	Account -ability	Minority inclusion
	(1)	(2)	(3)	(4)	(5)
e_van_comp	-0.002*	-0.002	-0.000	-0.001	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
van_comp × ethnic frac.	0.007***	0.006**	0.000	0.004**	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ln(gdppc)	-0.014	-0.002	0.000	0.008	0.010
	(0.03)	(0.03)	(0.01)	(0.02)	(0.01)
ln(population)	0.190	0.107	0.052	0.166*	0.030
	(0.12)	(0.10)	(0.04)	(0.08)	(0.02)
within R^2	0.0850	0.0720	0.0083	0.0524	0.0095
joint F-stat p-value	0.0165	0.0197	0.4346	0.1327	0.7549
Panel B		Share dependent variables Fixed effects OLS regressions			
		Policy learning	Electoral learning	Account -ability	Minority inclusion
		(1)	(2)	(3)	(4)
e_van_comp		-0.002	-0.000	-0.002*	-0.000
		(0.00)	(0.00)	(0.00)	(0.00)
van_comp × ethnic frac.		0.006**	0.000	0.005**	0.000
		(0.00)	(0.00)	(0.00)	(0.00)
ln(gdppc)		-0.002	0.002	-0.015	-0.000
		(0.03)	(0.00)	(0.02)	(0.00)
ln(population)		0.101	0.052	0.171*	0.013
		(0.10)	(0.04)	(0.09)	(0.01)
within R^2		0.0732	0.0090	0.0515	0.0092
joint F-stat p-value		0.0204	0.2739	0.0844	0.5407
Country & year FE's	yes	yes	yes	yes	yes
N	2718	2718	2718	2718	
Countries	160	160	160	160	160
N. experiments	286	255	46	95	23

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A14: Experiments and political competition across countries (Vanhanen 2016 measure)

Panel A	Binary dependent variables Fixed effects LPM regressions						
	All RCTs	Policy learning	Electoral learning	Political account	Bureau account	Decentral -ization	Minority inclusion
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
e_van_comp	-0.002*	-0.002	-0.000	-0.000	-0.001	-0.000	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
van_comp $\times$ ethnic frac.	0.007***	0.006**	0.000	0.001	0.003*	0.001	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
lngdppc	-0.014	-0.002	0.000	0.002	0.019	-0.016	0.010
	(0.03)	(0.03)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)
lnpop	0.190	0.107	0.052	0.057	0.070	0.078*	0.030
	(0.12)	(0.10)	(0.04)	(0.05)	(0.05)	(0.04)	(0.02)
within R^2	0.0850	0.0720	0.0083	0.0199	0.0407	0.0336	0.0095
joint F-stat p-value	0.0165	0.0197	0.4346	0.7693	0.0948	0.3993	0.7549
Panel B		Shares dependent variables Fixed effects OLS regressions					
		Policy learning	Electoral learning	Political account	Bureau account	Decentral -ization	Minority inclusion
		(1)	(2)	(3)	(4)	(5)	(6)
e_van_comp		-0.002	-0.000	-0.000	-0.001*	-0.000	-0.000
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
van_comp $\times$ ethnic frac.		0.006**	0.000	0.001	0.003*	0.002	0.000
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
ln(gdppc)		-0.002	0.002	-0.003	0.005	-0.016**	-0.000
		(0.03)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
ln(population)		0.101	0.052	0.059	0.041	0.071*	0.013
		(0.10)	(0.04)	(0.05)	(0.03)	(0.04)	(0.01)
within R^2		0.0732	0.0090	0.0182	0.0504	0.0324	0.0092
joint F-stat p-value		0.0204	0.2739	0.7727	0.1323	0.3650	0.5407
Country & year FE's	yes	yes	yes	yes	yes	yes	yes
N	2718	2718	2718	2718	2718	2718	2718
Countries	160	160	160	160	160	160	160

Notes: Robust standard errors clustered by country are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

C.3 Further results: Indian state data

Table A15: Experiments and political institutions in India

Panel A	Binary dependent variables Fixed effects LPM regressions				
	All RCTs	Policy learning	Electoral learning	Account -ability	Minority inclusion
	(1)	(2)	(3)	(4)	(5)
HH index of seat concentration	0.055	0.076	0.017	0.096*	-0.017
	(0.13)	(0.15)	(0.03)	(0.06)	(0.05)
ln(gsppc)	0.041	0.008	0.073	-0.029	0.110
	(0.17)	(0.16)	(0.09)	(0.11)	(0.09)
ln(population)	-0.083	-0.142	0.233	-0.104	0.128
	(0.40)	(0.35)	(0.23)	(0.21)	(0.23)
Panel B		Shares dependent variables Fixed effects OLS regressions			
		Policy learning	Electoral learning	Account -ability	Minority inclusion
		(1)	(2)	(3)	(4)
HH index of seat concentration		0.069	0.009	0.048*	-0.003
		(0.15)	(0.02)	(0.03)	(0.05)
ln(gsppc)		-0.018	0.038	-0.023	0.130
		(0.15)	(0.05)	(0.05)	(0.11)
ln(population)		-0.210	0.158	-0.113	0.192
		(0.33)	(0.16)	(0.14)	(0.27)
District & year FE's	yes	yes	yes	yes	yes
N	553	553	553	553	553
Districts	30	30	30	30	30

Notes: Robust standard errors clustered by Indian state are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A16: Experiments and political institutions in India

	Binary dependent variables							
	Fixed effects logit regressions				Random effects logit regressions			
	All RCTs	Policy learning	Electoral learning	Account -ability	All RCTs	Policy learning	Electoral learning	Account -ability
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HH index	2.134*	2.221**	6.704	3.964**	1.637	1.809*	-2.011	2.942*
	(1.12)	(1.11)	(12.36)	(1.81)	(1.05)	(1.04)	(1.82)	(1.61)
ln(gsppc)	-0.756	-0.656	-6.656*	-0.412	0.212	0.351*	-0.522	0.154
	(0.69)	(0.71)	(3.95)	(1.16)	(0.21)	(0.21)	(0.40)	(0.36)
ln(population)	8.872	9.050	55.567	7.424	1.605***	1.506***	1.747***	3.027**
	(5.67)	(5.88)	(39.12)	(10.29)	(0.39)	(0.36)	(0.56)	(1.33)
District FE's	yes	yes	yes	yes	no	no	no	no
N	266	266	40	137	553	553	553	553
Districts	14	14	2	7	30	30	30	30

Notes: Robust standard errors clustered by Indian state are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A17: Political accountability experiments and political competition in India

	Share accountability experiments Fixed effects OLS regressions						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HH index	0.044*	0.049*	0.048	0.054*	0.043	0.041	0.051*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
ln(gsppc)		-0.024	-0.012	-0.021	-0.015	-0.005	-0.012
		(0.05)	(0.06)	(0.06)	(0.07)	(0.06)	(0.08)
ln(population)		-0.113	-0.100	-0.168	-0.142	-0.083	-0.270
		(0.14)	(0.25)	(0.32)	(0.21)	(0.23)	(0.40)
state election		0.008					0.010
		(0.01)					(0.01)
tax revenues			0.002				0.003
			(0.01)				(0.01)
fiscal deficit				-0.000			-0.000
				(0.00)			(0.00)
literacy rate					0.380		0.500
					(0.51)		(0.57)
urbanization rate						0.001	0.002
						(0.00)	(0.00)
District & year FE's	yes	yes	yes	yes	yes	yes	yes
N	591	553	495	474	435	435	399
Districts	30	30	30	30	30	30	30

Notes: Robust standard errors clustered by Indian state are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.

Table A18: Experiments and political institutions in India

	Count dependent variables Random effects Poisson regressions				
	All RCTs	Policy learning	Electoral learning	Account -ability	Minority inclusion
	(1)	(2)	(3)	(4)	(5)
Margin seats	-1.128**	-1.076**	1.247	-2.019**	0.447
	(0.52)	(0.54)	(4.62)	(1.14)	(1.60)
ln(gsppc)	-0.101	0.005	-2.000	-0.106	-1.328***
	(0.12)	(0.13)	(1.52)	(0.32)	(0.49)
ln(population)	1.454***	1.400***	10.504	3.035**	4.563
	(0.32)	(0.32)	(11.54)	(1.42)	(3.16)
Random effects	yes	yes	no	yes	yes
N	553	553	553	553	553
Districts	30	30	30	30	30

Notes: Robust standard errors clustered by Indian state are in parentheses. *** / ** / * represent significance at the 1 / 5 / 10 percent level.