



# Managing performance and winning trust: how World Bank staff shape recipient performance

Mirko Heinzel<sup>1,2</sup> · Andrea Liese<sup>1</sup>

Accepted: 12 January 2021 / Published online: 2 March 2021  
© The Author(s) 2021

## Abstract

World Bank evaluations show that recipient performance varies substantially between different projects. Extant research has focused on country-level variables when explaining these variations. This article goes beyond country-level explanations and highlights the role of World Bank staff. We extend established arguments in the literature on compliance with the demands of International Organizations (IOs) and hypothesize that IO staff can shape recipient performance in three ways. First, recipient performance may be influenced by the quality of IO staff monitoring and supervision. Second, the leniency and stringency with which IO staff apply the aid agreement could improve recipient performance. Third, recipient performance may depend on whether IO staff can identify and mobilize supportive interlocutors through their networks in the recipient country. We test these arguments by linking a novel database on the tenure of World Bank task team leaders to projects evaluated between 1986 and 2020. The findings are consistent with the expectation that World Bank staff play an important role, but only in investment projects. There is substantial evidence that World Bank staff supervisory ability and country experience are linked to recipient performance in those projects. Less consistent evidence indicates that leniency could matter. These findings imply that World Bank staff play an important role in facilitating implementation of investment projects.

**Keywords** World Bank · International bureaucrats · Recipient performance · Enforcement · Supervision · Country experience · Compliance

---

*Responsible editor:* Christopher Kilby

---

✉ Mirko Heinzel  
mheinzel@uni-potsdam.de

Andrea Liese  
aliese@uni-potsdam.de

<sup>1</sup> University of Potsdam, August-Bebel-Straße 89, 14482 Potsdam, Germany

<sup>2</sup> Berlin Graduate School for Global and Transregional Studies, FU Berlin, Berlin, Germany

## 1 Introduction

Development finance agreements between the World Bank and recipient governments mandate a range of behaviors in implementation, including, for instance, commitment to the project, conducting stakeholder consultations, resolution of implementation issues, adequate financial management and adequate procurement. Data from the World Bank Independent Evaluation Group (IEG) evaluate the degree to which recipients fulfilled these obligations and “performed” well. The data show substantial variation between projects. Existing research explains these variations based on differences between countries (Girod and Tobin 2016). However, even when comparing projects in the same sector, from the same country, and approved in the same year, variation remains. Afghanistan, for instance, had two projects focusing on water and sanitation approved in 2005. In one, the IEG rated the performance as satisfactory, whereas in the other, it did not. This kind of variation is insufficiently understood in the literature on the effectiveness of development assistance and in broader debates on adherence to the agreements between recipients and International Organizations (IOs).

To explain variation in recipient performance, we build on studies that have shown that international institutions can differ in their service provision (Kilby 2013a; Smets et al. 2013) and complement this research by shifting the focus to the role of IO staff members. We emphasize that staff can foster recipient performance. The argument is built on a substantial body of literature focusing on the influence of international bureaucracies in IOs (Barnett and Finnemore 2004; Copelovitch 2010; Knill et al. 2019). These studies suggest that international bureaucrats have a considerable impact on policymaking at IOs in general (Barnett and Finnemore 2004; Hawkins et al. 2006; Knill et al. 2019) and the World Bank in particular (Honig 2020; Weaver 2008). Given that recipient performance can crucially hinge on the design or supervision of development finance agreements and development projects, the influence of staff members on recipient performance seems likely.

Therefore, we pose the following question: to what extent do international bureaucrats influence recipient performance in World Bank projects? Our argument is built on recent research on the World Bank that suggests that not all staff members execute their tasks in similar ways. Staff preferences (Briggs 2019a) and staff effectiveness (Bulman et al. 2017; Denizer et al. 2013) seem to differ substantially. We claim that these disparities matter for recipient performance. To explain variations, we hypothesize that three staff characteristics can influence recipient performance. These three characteristics are derived from theoretical perspectives in the IO literature that explain why member states comply with IO demands. We argue that international bureaucrats can be better at supervising projects, more stringent or lenient in applying development finance agreements, and more successful in securing the trust of essential stakeholders in recipient countries. Through these three pathways, international bureaucrats may directly influence the degree to which recipient states follow the aid agreement and, thus, perform well in the eyes of the Bank.

Our empirical results show substantial support for the hypotheses that World Bank staff characteristics matter for recipient performance. We demonstrate that recipient performance increases with the supervisory ability and country experience of staff. While we also show some evidence for the role of leniency, those findings depend much more on the variables employed and the specification used. Generally, the

evidence implies that staff play a critical role in facilitating the performance of recipients. However, the positive contributions of staff seem to materialize only for investment projects and not for development-policy operations. The findings are robust to extensive checks focusing on bias in ratings, employing alternative ways to measure the main independent variables, and employing a range of alternative model specifications. Furthermore, we probe whether self-selection might drive the results but find little evidence. Overall, evidence from a battery of different tests implies that World Bank staff are important in shaping recipient performance.

The article contributes to two debates in IO research: World Bank project performance and compliance with IO demands. Existing studies on World Bank project performance have primarily concentrated on project outcome ratings rather than recipient performance (Bulman et al. 2017; Denizer et al. 2013; Dreher et al. 2013; Honig 2019; Kilby 2015). They have demonstrated that the World Bank plays a substantial role in determining project outcomes: either through design (Smets et al. 2013) or through supervision (Kilby 2001). However, the literature on IOs more broadly has long argued that explanations focusing on the adherence to IO demands, often referred to as compliance (Börzel 2020; Girod and Tobin 2016), necessitate a different set of explanations than studies focusing on effectiveness (Simmons 1998). While research on the effectiveness of development assistance has concentrated on understanding the conditions under which development assistance leads to welfare gains, research on compliance has aimed to understand why member states adhere to IO demands in the absence of a centralized authority with coercive capacity (Hurd 1999). That said, whether welfare gains materialize through a project depends on much more than the adherence of recipients to the obligations spelled out in the aid agreement. Development impact might not materialize even if borrowers do everything that is expected of them, for example, because a project ignores essential local contexts. Critics of the Bank have often argued that Bank programs lack effectiveness or even do harm to development objectives (Easterly 2014). We deem it unlikely for an inappropriate project to become more effective, even when borrowers do what the World Bank expects of them. The crucial difference between outcomes and compliance is more than semantics. It has spurred debates that focus on explaining differences in performance and compliance, respectively, and emphasize different sets of variables (Simmons 1998). Our study explores recipient performance rather than project outcomes and complements the literature on World Bank project performance.

By doing so, we also contribute to the literature on compliance in International Relations. In their seminal study, Girod and Tobin (2016) contrast a strategic importance hypothesis with an alternative income hypothesis to convincingly demonstrate that countries with lower income dependence on World Bank projects are less likely to fulfil the expectations of the World Bank in the project. However, such differences between countries can still not account for a substantial degree of the variation we observe in recipient performance. This emphasis is also a feature of a larger body of the quantitative literature in International Relations (IR) that focuses on the adherence to an international agreement. This research on compliance focuses almost entirely on country-level variations (Börzel et al. 2010; Koliev et al. 2020; Simmons 2009; Vreeland 2006). We add to this compliance debate in two important ways: first, we extend its arguments with insights on the characteristics of international bureaucrats

from the respective literature in public administration. Second, we demonstrate empirically that the variation in adherence to World Bank assistance agreements can be explained by variations between the international bureaucrats tasked with overseeing the project.

We proceed in five steps. First, we discuss why we expect staff to influence recipient performance. Second, we extend well-established perspectives on compliance from international relations and develop three hypotheses on how World Bank staff can shape the stringency of enforcement, build the capacity of recipients, and foster the sympathy of interlocutors. Third, we introduce the data used to test these hypotheses. Fourth, we present evidence to evaluate the relative importance of country and staff level factors in explaining recipient performance. Finally, we conclude by discussing the implications of our findings for academic debates on both World Bank effectiveness and country adherence to the rules of the agreement with an IO.

## 2 Recipient performance and the role of task team leaders

Recipient performance in the context of World Bank projects can mean different things depending on different types of projects. There are two main types of lending instruments the World Bank uses: policy-based lending and investment project loans.<sup>1</sup> Both categories of operations include an agreement that stipulates the obligations of borrowers in the context of the loan. For Development Policy Financing (DPF), the contract includes conditionality that borrowers are supposed to implement. The particular set-up has changed over time from Structural Adjustment Loans (SALs), where conditionality was implemented in exchange for disbursements, to Development Policy Loans (DPL), where the implementation of conditionality precedes loan approval.<sup>2</sup> Investment Project Financing (IPF), on the other hand, is tied to specific projects. Here, the World Bank commits money in the expectation that it is used for a particular project that was previously agreed upon with a recipient. The project is implemented by the borrowing government and implementing agencies in the recipient country. Recipient adherence to the aid agreement contains less precise conditions to be implemented and more generally ascribed behaviors in IPF. These behaviors include, for example, government ownership, stakeholder consultations, resolution of implementation issues, as well as adhering to adequate financial management, procurement guidelines and counterpart commitment.

We argue that World Bank staff can play an important, underappreciated role in facilitating recipient performance in World Bank projects. Specifically, we focus on one group of staff members: Task Team Leaders (TTL). They are the most relevant staff for individual projects (Briggs 2019a; Bulman et al. 2017; Denizer et al. 2013), although several different staff at different levels of the hierarchy are also involved in

<sup>1</sup> In the interest of parsimony, we speak of lending throughout the article even though some of the loans are partly or fully given as grants. Recent years have seen a rise in program for results funding, which could be classified as a third main type of lending types (Cormier 2016). However, this shift had not taken effect yet in the period under study.

<sup>2</sup> We refer to DPF when we talk about development-policy financing regardless of whether the specific kind of DPF the World Bank administered was called SAL, DPL, or another credit line was used that is subsumed under DPF.

projects. Most notably, all work with recipients is managed by the country management unit and headed by the country director. In large countries, like India, the World Bank has one country office exclusively focusing on this country. In other cases, country directors are in charge of multiple countries, like the country director for Kenya, Rwanda, Eritrea and Uganda. Country directors work with the recipient government and World Bank sectoral staff to converse on project ideas. Once the general idea has been determined, a TTL is chosen who works as “the Bank’s principal point of contact for the borrower for the project” (World Bank 2013a, p. 1). TTLs are usually recruited by the practice managers responsible for a project. The project team is headed by the TTL and further includes technical staff working on legal issues, financial management and safeguards.

TTLs have a substantial role to play both in the preparation and implementation of projects (World Bank 2013a). In the preparation of a project, TTLs lead several missions in the recipient country. These missions include the identification mission, the pre-appraisal mission, and the appraisal mission. TTLs are also in charge of drafting all relevant documents while designing the project. Such documents are the project information documents, the project concept notes and project appraisal documents. Thereby, TTLs spearhead planning on procurement, financial management, project objectives, risk identification, and project instruments. After the initial planning phase, TTLs head the negotiation team of the World Bank in the final negotiations with the recipient. They prepare those documents that the executive board uses when approving projects.<sup>3</sup>

During implementation, World Bank projects are in the hands of the recipient government and implementing agencies. Nevertheless, TTLs play a fundamental role. They lead the World Bank’s implementation support to the recipient government and are in charge of producing implementation reviews that are updated at a minimum every six months. These reviews document the progress the project has made toward its objectives and are also used as the information basis for the task team when deciding where to focus efforts, whether to reduce the resources available in the project or to propose restructuring a project. Furthermore, TTLs are in charge of producing reviews that are conducted roughly at the mid-point of the project. These mid-term reviews can lead to changes in closing dates, alternations to the targets of the project as well as plans on restructuring. The TTL leads decisions on proposing a restructuring of a project to the executive board and prepares all relevant documents for meetings discussing restructuring. Additionally, the TTL is in charge of identifying governance and corruption issues in the project. TTLs review selected contracts, procurements and financial management to identify these issues.

In a limited number of major decisions, TTLs do not have a meaningful say. Formally, they are not authorized to decide on additional financing requests (decided by the country director) or final decisions to submit the project for approval to the executive board (decided by the regional Vice President) and subsequent approval (decided by the executive board). However, TTLs are involved in drafting all documents that are the basis for the decisions of these actors (World Bank 2013a). Due to their wide-ranging responsibility throughout the project-cycle, they can be seen as the most crucial World Bank staff members for each project (Briggs 2019a; Bulman et al. 2017; Denizer et al. 2013).

<sup>3</sup> TTLs are not present at board meetings, where the project is presented by regional vice-presidents

### 3 Three explanations for the influence of staff

After having discussed TTLs centrality for World Bank projects in general, we derive hypotheses regarding variations between different TTLs in their ability to foster recipient performance in World Bank projects. To do so, we need to make assumptions about the actions of individual staff members and on the actions of borrowers in a project. We assume that individuals show consistency in their actions and that there are variations between individuals working in the same organization. The assumption is based on a substantial body of literature in psychology and public administration research that focuses on individual learning, routines and decision-making by “street-level” bureaucrats (Eckhard 2020; Lipsky 1980; May and Winter 2009). Furthermore, the assumption builds on the evidence presented by Denizer et al. (2013) on World Bank TTLs. These authors show that TTL-fixed effects explain a sizeable part of the variation in project outcomes. We present similar evidence for recipient performance below. Our findings imply consistency in the decisions of individuals across contexts as well as variations between individual staff members. However, which variations matter when explaining recipient performance?

To ascertain relevant TTL characteristics, we build on the burgeoning literature on compliance with international agreements in International Relations (Raustiala and Slaughter 2002; Simmons 1998; Tallberg 2002). It has focused on a central problem all social systems confront: “the problem of social control – that is, how to get actors to comply with society’s rules” (Hurd 1999, p. 379). This problem is particularly pronounced in global governance because IOs lack the extent of coercive capacity that national governments commonly possess. In the case of the World Bank, the problem of social control materializes in the necessity to ensure that recipient governments do not just “take the money and run” (Girod and Tobin 2016, p. 209). The Bank wants to assure that recipients adhere to the aid agreement and perform to its satisfaction. Notably, we derive expectations from three central perspectives that make different assumptions on the reasons why borrowing governments might be lacking in recipient performance. These three perspectives focus on either the ability or the willingness of actors in the recipient country to adhere to the aid agreement. We identify staff characteristics that would conceivably allow staff to make positive contributions. The derived hypotheses follow on building either the capacity (supervision) or the will (leniency, country experience) of recipients to perform well in the view of the Bank. We discuss the three perspectives in turn.

#### 3.1 Supervisory ability and recipient performance

The first perspective, commonly referred to as the management school of compliance research, explains recipient performance based on the cost-benefit calculations of rational recipients. Non-adherence to a commitment is a feature of lacking capacity or imprecise rules (Chayes and Chayes 1993). This perspective corresponds to what has been called a “cooperative model” of World Bank-recipient relations in the World Bank literature (Kilby 2001, p. 205). The World Bank and the recipient have common interests, and any failure to “perform” on the part of the recipient must be due to technical concerns. Indeed, a range of issues during the implementation of a project can undermine the ability of a recipient government and its implementing agencies to

perform well. The literature on the effectiveness of development assistance has emphasized corruption and state fragility as a central inhibitor for the success of development projects (Dollar and Levin 2005; Dollar and Svensson 2000; Honig 2019). If lower-level officials in the government or the implementing agencies work to enrich themselves through the project, satisfactory performance is substantially less likely. The recipient government will have agreed to a project that it plans to implement, but it will not have the capacity to induce its representatives to do so. Donors react to the level of corruption in recipient countries and try to minimize the use of country implementation systems in specific contexts (Dietrich 2013; Knack 2013; Winters 2014). Similarly, if a country is faced with substantial fragility, the project environment is much less predictable, which necessitates more decentralized decision-making in implementation (Honig 2019; Marchesi and Masi 2020).

We expect some TTLs to be better able to support recipients through better supervision. For example, former World Bank staff member Paul Cadario directly discusses differences between staff in that regard: “with the exception of some of the weaker loan officers, people knew the country, the managers were very hands-on, they rolled their sleeves up, they read things, they gave you feedback (...), they edited things if they didn’t have the right direction on them” (World Bank 2013b, p. 7). Supervision entails monitoring of project implementation and technical assistance to the recipient (Kilby 2001). Better supervision allows for earlier identification of implementation issues and better ability to remedy these problems. The efforts staff exert in supervision can vary. Authors have convincingly demonstrated that country-level factors like government ideology or geopolitical alignment to the World Bank’s major shareholders can explain why efforts vary across staff (Smets et al. 2013; Wane 2004). We argue that individual differences between staff members also explain why the World Bank will vary in its quality of supervision. There are differences between TTLs in the importance they ascribe to project effectiveness (Briggs 2019a) and also their competence (Denizer et al. 2013; Limodio 2021). These differences shape the ability of staff to build capacity in recipient countries and help willing governments to adhere to the World Bank’s expectations. The first hypothesis hence reads:

H1: The better a given international bureaucrat is at supervising a project, the more likely the recipient is to comply.

### 3.2 Leniency and recipient performance

The second perspective, which is often referred to as the enforcement school in the compliance literature, also argues that recipient actors make decisions based on cost-benefit calculations but shifts the focus to willingness rather than capacity (Downs 1998; Fearon 1998). In this view, recipients follow IO demands if enforcement inflates the costs of not doing so. This perspective has been called the “adversarial model” in the World Bank literature (Kilby 2001, p. 203). The argument is that recipient governments could implement as little as possible to retain disbursement while lacking real commitment to the project. Here, lacking commitment at the higher levels of the recipient government translates to less effort in the implementing agencies. This is the perspective Girod and Tobin (2016) take in their study of recipient performance. They

argue that governments that rely substantially on World Bank assistance for their budgets and public services will be more committed to adhering to the aid agreement. On the other hand, those governments that have alternative sources of revenue, like resource rents or FDI, will be much less interested in diligent implementation (Girod and Tobin 2016). The second variant of enforcement approaches is presented by Stone (2002) based on an analysis of the IMF's role in economic transitions of several Eastern European countries. He argues that member states that are strategically important to the organization and its major shareholders will be able to secure disbursements even in the absence of compliance with IMF conditionality. Indeed, evidence on the World Bank suggests that strategically important countries seem to be able to secure more disbursements (Andersen et al. 2006; Kilby 2013b). Recent evidence, on project design rather than disbursement, suggests that preferential treatment might not occur because of the lobbying of powerful member states but rather because staff either internalize shared preferences or subconsciously try to "please the principal" (Clark and Dolan 2021, p. 4). Whether powerful shareholders lobby staff directly or whether staff internalize considering their preferences, differences in disbursement might translate into lacking effort on the part of the recipient government.

We argue that bureaucratic preferences also play a decisive role in enforcement throughout the project. World Bank bureaucrats have discretion in enforcement decisions because staff are central for decisions to release disbursements. However, TTLs are caught between two objectives. On the one hand, they are expected to contribute to the effectiveness of projects positively. On the other hand, TTLs prestige and career advancement depends on the disbursement of projects and getting projects with high loan volumes approved (Briggs 2019a; Weaver 2008). As one report by the Independent Evaluation Group highlights, staff have the "perception that individual success depends more on obtaining new deals and ensuring timely disbursement than on quality implementation" (World Bank 2016, p. 28). TTLs differ in how they square these two, at times competing, objectives. Recent survey evidence presented by Briggs (2019a) including responses by 115 TTLs indicates that around half of TTLs think that project ratings are not at all or only slightly important for their career success, while ca. 25% think they are extremely or very important. Therefore, there are likely variations between staff in how stringent they are in basing disbursement decisions on the implementation record of the project. Anecdotal evidence illustrates this argument. Former World Bank staffer Jan Walliser describes the work of another staff member in his interview with the oral history project: "We went out to a site that had been identified for a multi-donor project but where construction was only about to begin. At that point, the project had actually encountered quite a bit of trouble. And I will never forget how she stood on the sides of the river there, interrogating the Russian contractor who had not properly mobilized the equipment to start the works despite some of the prepayments that had been made, and questioning him trying to get some answers on how the project could get off the ground" (World Bank 2018, p. 40). Given the incentives in the Bank, not all staff members show this same dedication to basing disbursement decisions on the efforts exerted by recipients. If recipients discover that a TTL is less stringent on gathering information on lacking implementation and aims to ensure disbursements regardless, they will start lagging behind. The corresponding hypothesis is:



H2: The more stringent an international bureaucrat is in enforcement, the more likely the recipient is to comply.

### 3.3 Country experience, sympathetic interlocutors, and recipient performance

The third perspective focuses on the will of recipient actors as well but shifts the focus from cost-benefit calculations to the recipients' receptivity for World Bank advice. It is derived from ideational perspectives on compliance (Checkel 1997; Simmons 2013). In the literature on International Financial Institutions, these perspectives have been discussed under the label of sympathetic interlocutors (Arpac and Bird 2009; Bazbauers 2019; Heinzel et al. 2020; Woods 2006). According to Woods (2006, p. 10), sympathetic interlocutors are understood as actors that are "both willing and able to embrace the priorities preferred by the institutions." A lack of sympathetic interlocutors can impede the national implementation of IO policies (Broome and Seabrooke 2015). Woods (2006) argues that the International Financial Institutions need to identify existing sympathetic interlocutors inside the national bureaucracy or persuade interlocutors to increase their commitment to reform. These two means of cultivation are further specified by Bazbauers (2019). To identify sympathetic interlocutors, the World Bank staff can either locate them through their public support for programs or based on their relationships built through prior engagements. Alternatively, World Bank staff can seek to persuade interlocutors of the importance of the project. Persuasion can work through direct interaction or by mobilizing already existing interlocutors (Bazbauers 2019).

Country experience crucially influences the ability of IO staff to identify sympathetic interlocutors and persuade recipients to embrace the World Bank's priorities. First, identification of sympathetic interlocutors in a recipient country is eased if staff have prior experience in this country. Experienced staff likely have a better sense of the actors in the national bureaucracy and implementing agencies that share the objectives of a project. Their networks in the recipient country will also allow them to seek out sympathetic interlocutors and try to work with them on the project. That way, their country experience will equip them with more of the necessary tools for the identification of people that may help steer the domestic implementation process. Second, persuasion of opponents of projects is also eased when staff have country experience and, thus, can utilize their existing personal and professional relationships to foster support for the new project. This insight is built on findings in social psychology and organizational studies that indicate that more prolonged exposure to individuals increases trust, *ceteris paribus* (Bomstein 1989; Kwan et al. 2015). Several studies on development projects have pointed to the importance of trust and communication between national and international implementation staff. These authors argue that when recipients trust World Bank staff, they will be more open to dialogue and persuasion, and be more inclined to adhere to their obligation in a project (Bazbauers 2019; Ika et al. 2012; Lannon and Walsh 2020). Indeed, research on social trust shows that trust renders policy recommendations more credible and leads to actors being more receptive to external advice and criticism (Keefer and Knack 2008). These factors, in turn, render persuasion more likely. For example, former World Bank staffer Jan Walliser discusses his experience of working with country counterparts: "I developed very close relationships with the counterparts, and the sense of trust, especially with the Finance Minister (...). And they were largely built on, that they knew or that they got over time, the sense that what I

was saying was to help them rather than to make a point” (World Bank 2018, p. 10). This is not to say that all relationships between World Bank staff and interlocutors will be positive. Government counterparts do not automatically listen to staff just because they have worked in the country before. However, staff with country experience even have an advantage when prior experiences were negative, and a prior World Bank project was met with resistance. In these cases, they will have a better sense of who might not be receptive to their advice than those who lack country experience. Therefore, staff with country experience have an advantage over staff without country experience when trying to both identify and persuade interlocutors. Based on these considerations, the third hypothesis is:

H3: The more country experience a given international bureaucrat has, the more likely the recipient is to comply.

## 4 Research design

To assess the three developed hypotheses, we built a database of World Bank projects. It combines novel data collected on the main staff members in charge of the implementation of projects with available data on projects from the oftentimes used World Bank projects database, Independent Evaluation Group (IEG) evaluation data and country-level variables (World Bank 2020a). We discuss the dependent variable, staff level independent variables and country-level independent variables in turn.

### 4.1 Dependent variable

When measuring recipient performance, we draw on project evaluations produced by the IEG between 1986 and 2020. The IEG is the standard data source for analyzing World Bank project performance (Bulman et al. 2017; Denizer et al. 2013; Dreher et al. 2013; Kilby and Michaelowa 2019; Winters 2019) and recipient performance (Girod and Tobin 2016). At the end of a project, the World Bank staff evaluate each project based on a predefined set of criteria. The IEG desk reviews the projects to ensure consistency. Projects are sometimes re-evaluated by IEG. When multiple ratings are available for the same project, we use the most recent rating.<sup>4</sup> Projects are rated through two kinds of reports: Implementation Completion Reports (ICR) and Project Performance Assessment Reports (PPAR).<sup>5</sup> PPARs make up around 30% of evaluations.<sup>6</sup>

<sup>4</sup> We use the most recent available rating for all variables drawing on IEG data throughout the article.

<sup>5</sup> 21 of the earliest projects in the database are rated through Project Completion Reports. Project Completion Reports were issued between 1982 and 1994. They were replaced by ICR in 1995.

<sup>6</sup> Kilby and Michaelowa (2019) provide a detailed discussion of IEG’s selection of projects for PPARs. They find that many of the patterns of selection are consistent with IEG’s mission. The IEG is more likely to choose very positive evaluations and evaluations where the ICR quality is rated as very poor. Projects with larger loan volumes are selected more often and IEG selects investment projects less frequently than program loans. However, they also show that PPARs are more often in countries where working conditions are easier and that are popular tourist locations. Finally, they find evidence that IEG seems to consider institutional power when determining the release date of the reports and that a seat in the UNSC can have an impact on the PPAR rating. These findings show that PPARs might not be as independent in general but they should still be more independent from the TTL than the ICRs (and we provide evidence that staff variables are not significant predictors of rating revisions in Appendix A5).

They include both extensive review of World Bank documentation and visits to the recipient country to verify ratings in detail. In about one in five evaluations, IEG adjusts the ratings because of disagreements with the decisions of initial evaluators (Malik and Stone 2018).

We draw on IEG data that measure “the extent to which the borrower (including the government and implementing agency or agencies) ensured quality of preparation and implementation, and complied with covenants and agreements” (World Bank 2015b, p. 21). The IEG rates recipient performance based on two criteria: the performance of the borrowing government and the performance of the implementing agency in member states. Ratings are produced on a scale of 1–6 from highly unsatisfactory (1) to highly satisfactory (6). Figure 1 displays the distribution of the recipient performance evaluations. As is the case with all IEG evaluations, they skew rather positively. We perform some robustness checks to account for the overall positive ratings (see below).

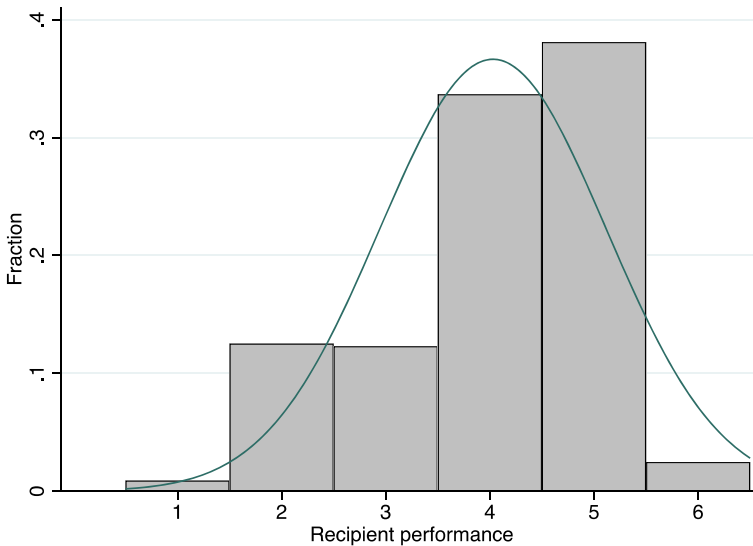
## 4.2 Staff-level variables

To operationalize the three hypotheses on the influence of World Bank bureaucrats, we draw upon a novel database on World Bank TTLs in charge of World Bank projects. We collected data on TTLs from the World Bank website. We scraped the names of the leading individuals, in charge of projects at implementation, using the publicly available World Bank API. We have data for 10,000 of the roughly 18,000 projects the World Bank has run in its history.<sup>7</sup>

The data present us with two challenges when calculating the indicators: first, increasing coverage and second, rotation of staff. The data set has excellent coverage after 2010 (more than 90%), and for half of the projects in the 1990s (for details, see Fig. A2). However, it is unlikely that staff-specific factors influence the likelihood that the World Bank gives details on a particular project on its website. Instead, reporting is likely influenced by developments in technology and increased transparency over time. The pattern of missingness is not substantively related to observed rating quality, outcome or recipient performance ratings or project amounts (Figs. A3–A4).

In addition, staff rotation might be an issue for constructing the variables. The World Bank website records the TTL at the end of the project rather than the one that started the project. TTLs tend to rotate every 3–5 years (Denizer et al. 2013; Kilby 2013b). This rotation pattern is based on the so-called 3–5–7 rule of the World Bank. The rule mandates being in assignment for a minimum of three years. After five years, staff are encouraged to pursue reassignment to another vice-presidential unit. When seven years have elapsed, regional managers actively facilitate the rotation of staff (World Bank 2015a, p. 62). To contextualize our

<sup>7</sup> The staff variables we use are based on past behavior and experiences of TTLs. This approach is built on Denizer et al. (2013) and Bulman et al. (2017) who estimate “TTL quality” based on the past outcome ratings of projects. We believe that we can get a reasonable estimation of differences between TTLs by drawing on their past record, because authors working on public administration have long argued that staff develop routines throughout their career that make their actions more consistent (Lipsky 1980). Nevertheless, our approach has the limitation that we ascribe project ratings to the past behavior and experiences of TTLs that they might not be the sole reason for.



**Fig. 1** Distribution of IEG recipient performance ratings. Note: The figure is based on the latest IEG recipient performance rating available in the database (3562 projects)

analysis, we randomly selected 200 projects and collected data on the TTLs in charge of appraisal and at the time of the ICR from the ICR documents. Table 1 displays patterns of rotation in these 200 projects. The vast majority of projects sees a change of TTLs from appraisal to implementation. This applies to both DPLs and IPF. The average length of those projects that share TTLs at appraisal and implementation is around 500 days shorter than when TTLs change. Substantially fewer TTLs change between implementation and evaluation. While we cannot account for the patterns of rotation directly, we try to rectify these issues in the construction of our supervision variable by only drawing upon the supervision component of the indicator and by controlling for project length in a robustness check.

One can reasonably disagree with a range of choices we made when constructing the indicators. To account for such disagreements, we present several alternatives to calculate our variables in the appendix (Tables A8-A10). We now discuss the three variables we use in the main body of the article. If more than one TTL is listed for one

**Table 1** TTL rotation based on a random sample of 200 projects in the database

	Shared TTL at appraisal and implementation	Different TTL at appraisal and implementation	Shared TTL at implementation and ICR	Different TTL at implementation and ICR
Total ( $n=200$ )	7.5%	92.5%	79.0%	21.0%
DPL ( $n=26$ )	11.5%	88.6%	96.0%	4.0%
IPF ( $n=174$ )	6.8%	93.2%	76.5%	23.5%
Average length	1942 days	2459 days	2406 days	2546 days

project, we take the average value of each of the variables for each TTL listed on a given project. We do so for all three variables.<sup>8</sup>

First, we measured the past performance of a given bureaucrat's projects to operationalize H1 on staff *supervision*. In constructing this indicator, we assume that past behavior is a predictor of future behavior. IEG evaluates the degree to which a project was designed well (quality at entry) and the degree to which staff helped recipient governments to identify and rectify issues in implementation (quality of supervision). We draw on supervision alone because our data do not include the TTL that was in charge of design for most projects (see above). To calculate the supervision indicator, we take the average supervision score for all projects that a given TTL was in charge of before the project of interest was approved (ranges from 1 to 6).

Second, to test H2, we constructed an indicator for the *leniency* of staff. The operationalization of TTL leniency draws on the observed leniency in TTL's past projects. The reasoning is that TTLs that were lenient in the past are more likely to be lenient in the future. We use a measure of disbursement share conditional on recipient performance in projects. The disbursement indicator is calculated by dividing total disbursement by total commitment of each project according to the World Bank (2020a). We follow Malik and Stone (2018) in truncating disbursement share data between 0 and 1. To build the leniency indicator, we divide the disbursement of projects (range from 0 to 1) by their recipient performance rating (range from 1 to 6).

Third, to operationalize H3 focusing on the trust of interlocutors, we measure the degree to which someone is embedded in local networks. We argue that TTLs that have worked previously in a given country and built substantial networks are better able to cultivate sympathetic interlocutors. We use the number of projects someone has run in a given country as an approximation of the integration into country networks. To do so, we draw on all projects listed in the World Bank projects database for which we have TTL data (10,196 projects).<sup>9</sup> We then take the natural logarithm (+1) to account for the decreasing marginal returns of experience (Yelle 1979).<sup>10</sup> The equations used to calculate all three staff variables can be found in the appendix (pages 2–4). One limitation of this approach is that the experience of TTLs that are newly hired is more fully captured than of those of TTLs which have been working for the World Bank much longer.<sup>11</sup>

<sup>8</sup> These co-TTLs are teams of TTLs that work on one project together. Co-TTLs occur in the larger database on TTLs roughly 19% of the time. However, two or more TTLs work only on roughly 3% of the projects in the database. This is for two reasons. First, co-TTLs are much more likely for projects focusing on multiple countries or whole regions. These projects are excluded from the analysis here, to not artificially create country-level variables from projects spanning multiple countries. Second, co-TTLs have become more numerous over time. Since many of the most recent projects have not been evaluated (see Figure A2), these projects are not in the database. When multiple TTLs are staffed, we average staff variables for one project. We further include a dummy variable controlling for the presence of co-TTLs in all models.

<sup>9</sup> They include IDA, IBRD and World Bank trust fund projects.

<sup>10</sup> We add +1 to account for projects where experience is 0.

<sup>11</sup> The earlier years are more impacted than the latter years by the truncated records of TTL appointments. When looking at the whole TTL database, the mean number of projects the TTL has supervised in a given country is 0.19 in the 1980s (SD 0.588), 0.32 in the 1990s (SD 0.76), 0.43 in the 2000s (SD 0.93) and 0.91 in the 2010s (SD 1.58). While only circa 13% are indicated as having worked in the country before approval of a given project for projects approved in the 1980s, roughly 46% of TTLs have observed country experience in the 2010s.

### 4.3 Control variables

We use control variables on the country- and project-level to minimize the possibility that our findings are driven by unobserved heterogeneity. When choosing control variables, we aim to account for variations the three perspectives on adherence to IO demands would deem crucial. All variables are measured in the year the project was approved unless explicitly discussed otherwise.

We employ a range of country-level variables to account for the different channels affecting recipient performance on the country-level. First, we control for variations in the strategic importance of recipient countries. To identify the necessary control variables, we can draw on longstanding debates on the informal influence of major shareholders at the World Bank (Stone 2013). Qualitative case studies (Babb 2009; Wade 2002; Woods 2006), as well as quantitative analyses (Andersen et al. 2006; Clark and Dolan 2021; Vreeland 2019), have demonstrated that US allies often get a better deal from the World Bank. In line with recent research on the World Bank (Clark and Dolan 2021), we use the distance between ideal points calculated based on important UNGA votes to measure such geopolitical alignment (Bailey et al. 2017). Second, credibility in the application of development finance agreements hinges on the power of the recipient country (Börzel et al. 2010; Stone 2002). It is more difficult to pressure powerful recipients like China compared to small island states like Kiribati. In line with research on rule adherence, we include a country's GDP, from the World Bank World Development Indicators (World Bank 2020b), as a predictor to account for these variations (Börzel et al. 2010). Furthermore, GDP is also crucial because it can be seen as a measure of state capacity. Third, Girod and Tobin (2016) have argued that it is more challenging to enforce development finance agreements when countries have alternative sources of income, like FDI. Therefore, we include data on FDI inflows from the World Development Indicators (World Bank 2020b). Fourth, we include control of corruption as an indicator of capacity constraints. The World Bank has often stated that implementation of its projects is hindered by governance issues in recipient countries (Aguilar et al. 2010; Wolfensohn 1996) and the argument has been prominently featured in the literature on the effectiveness of development assistance (Dollar and Levin 2005; Knack 2013; Winters 2014). We use the Varieties of Democracy Project's (VDem) public corruption indicator to measure country-level corruption (Coppedge et al. 2019). Finally, we control for regime type in the recipient country to account for varying size of the electorate of policymakers in recipient countries that can shape the incentives of governments to adjust policy (Bueno De Mesquita and Smith 2009). To do so, we employ VDem's polyarchy index.

In addition to these country-level variables, we control for differences between projects. First, we include the (log) project amount to take differences between the relevance of projects into consideration. Some projects include higher monetary commitments than others. If these projects are not implemented diligently and the World Bank does not disburse, the recipient country would lose more money than in small projects (Girod and Tobin 2016; Smets et al. 2013). Data on project amounts are taken from the IEG evaluation database (World Bank 2020a). Second, we try to account for heterogeneity in administrative procedures of the World Bank by including a dummy indicating whether projects were DPF. To control for variation between recipients and projects that the discussed variables cannot account for, we further employ country-

approval-year and sector-approval-year fixed effects in some specifications. Third, we operationalize arguments regarding the fit of international demands with government preferences by creating a dummy on whether the government changed during a project. The rationale is that government are more likely to take ownership of and thus be committed to the successful implementation of projects they initiated (as compared to projects a previous government initiated). Data on government change comes from the IADB's Database of Political Institutions (Scartascini et al. 2017). Finally, we control for projects supervised by multiple TTLs by employing a dummy variable that indicates whether the project was supervised by co-TTLs.

## 5 Analysis

We proceed in three steps to assess our argument that staff can influence recipient performance. First, we discuss the relative importance of country-level and staff-level variation in explaining the performance of recipients in the eyes of the World Bank. We do so to support our general argument that staff have a substantial impact on recipient performance. Second, we analyze the relative explanatory power of the three theory-driven explanations on the staff level and assess their magnitude in comparison to commonly discussed country-level variables. Third, we compare different loan types and actors on the national level that are involved in World Bank projects.

### 5.1 The relative importance of recipient country and World Bank staff in recipient performance

The basic premise of this article is that the literature has not paid sufficient attention to the role of staff in fostering adherence of recipients to development finance agreements. To probe that argument, we first discuss the relative importance of staff-level compared to country-level explanations.

We are not the first to ask the question of how much staff matter in World Bank projects. Indeed, Denizer et al. (2013) present evidence that World Bank staff can positively influence the outcome of World Bank projects. They show that the influence of World Bank staff is about as pronounced as all country factors combined. While their findings are highly relevant, it is not clear whether the same applies to recipient performance. Therefore, we replicate their analysis regarding the comparison of country and staff-level factors in Table 2. To do so, we limit the analysis to those staff

**Table 2** Analysis of variance of recipient performance ( $r^2 = 0.477$ )

	Partial sum of squares	Degrees of freedom	Mean sum of squares	F-statistic	Prob > F
TTL fixed effects	1216.45	952	1.28	1.31	0.000
Country fixed effects	188.18	141	1.33	1.36	0.004
Residual	1790.65	1831	0.98		
Total	3421.30	2924	1.17		

members who were responsible for at least two projects in at least two countries in our database. We then use a simple two-way ANOVA to separate variation in recipient performance into variation due to staff and country fixed effects (Denizer et al. 2013).

The model accounts for 48% of the variation in recipient performance ratings. Both country and staff fixed effects are jointly highly significantly different than zero. The mean sum of squares allows for a comparison of the relative explanatory power of country and staff factors. It accounts for the different number of fixed effects of both factors, which means that the sheer higher number of TTL fixed effects does not drive the results. The mean sum of squares of staff factors is on a similar order of magnitude as country factors. Therefore, it can be said that the staff seem to have a substantial impact on recipient performance. The strength of the finding increases the importance of better understanding what drives the differences in staffs' ability to contribute to recipient performance positively.

## 5.2 Assessing the three explanations for staff influence

We assess the three hypotheses focusing on the contribution of staff to recipient performance in turn.<sup>12</sup> All models used in the main body of the article are OLS regressions<sup>13</sup> with standard errors clustered at the evaluation-year.<sup>14</sup> Our estimation approach allows for correlated errors due to similar IEG evaluation practices in the same financial year (Denizer et al. 2013). Descriptive statistics can be found in the appendix (Table A2). In Table 3, we present the results for the argument that staffs' supervisory capacity influences recipient performance (H1). In the first three presented regressions, we alter the model specification by varying fixed effects to account for different sources of unobserved heterogeneity. Model 1 includes project-level and country-level controls as well as country, sector and approval-year fixed effects. While we account for a range of possible confounders on the country-level directly, we might still face omitted variable bias in particular country-approval-years or sector-approval-years. Therefore, we include the respective fixed effects in Models 2 and 3. Finally, fostering recipient performance includes both higher-level political engagement with the government as well as lower-level administrative coordination with implementing agencies. We assess whether the explanatory potential of staff supervision differs between government performance and implementing agency performance in Models 4 and 5.

The findings presented in Table 3 indicate that the supervisory ability of TTLs can affect recipient performance. TTL supervision is a positive and statistically significant predictor of performance ratings in four of the five models. Even when accounting for country-approval-year (Model 2) and sector-approval-year fixed effects (Model 3), the coefficient is significant ( $p < 0.1$ ). However, when disaggregating by government and recipient performance, we find that supervision seems to work more consistently

<sup>12</sup> The sample size of models differs throughout because the staff supervision and leniency indicators draw upon the past project records of TTLs. When TTLs run their first project, such a record is not available and they are omitted from the sample. Furthermore, while we have data on TTLs records for 10,196 projects, recipient performance evaluations are only available for 3,562 of these projects.

<sup>13</sup> To increase computational efficiency we use the algorithm developed by Guimaraes and Portugal (2010) to fit models with high-dimensional fixed effects.

<sup>14</sup> When discussing evaluation-years or approval-years throughout, we refer to financial years rather than calendar years.



**Table 3** Results of OLS models regressing recipient performance rankings on supervision

	(1) Recipient	(2) Recipient	(3) Recipient	(4) Government	(5) Agency
Staff supervision	0.0806* (0.0424)	0.1152* (0.0664)	0.0790* (0.0389)	0.1078 (0.0682)	0.1230* (0.0643)
Project-level controls	Yes	Yes	Yes	Yes	Yes
Country-level controls	Yes	No	Yes	No	No
Country fixed effects	Yes	No	Yes	No	No
Country-approval-year fixed effects	No	Yes	No	Yes	Yes
Sector fixed effects	Yes	Yes	No	Yes	Yes
Sector-approval-year fixed effects	No	No	Yes	No	No
Approval-year fixed effects	Yes	No	No	No	No
Observations	1766	1136	1727	1112	1077
R <sup>2</sup>	0.199	0.487	0.327	0.451	0.446

Note: Standard errors in parentheses; errors clustered at the evaluation-financial-year; project-level controls: DPF dummy, project amount (log), government change throughout the project and co-TTL dummy; country-level controls: corruption, ideal point (based on votes the US state department designates important) difference between recipient and US, GDP (log), FDI and democracy; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

through agency performance. While the coefficient is on a similar order of magnitude for both groups of recipient country actors, it fails to attain statistical significance for government performance (Model 4).

In a second step, we evaluate the association between the leniency shown in past projects of a given TTL and recipient performance. The results are displayed in Table 4.

**Table 4** Results of OLS models regressing recipient performance rankings on leniency

	(6) Recipient	(7) Recipient	(8) Recipient	(9) Government	(10) Agency
Staff leniency	-0.6880* (0.3632)	-0.7340 (0.5811)	-0.7843** (0.3783)	-0.2436 (0.6281)	-0.2171 (0.7160)
Project-level controls	Yes	Yes	Yes	Yes	Yes
Country-level controls	Yes	No	Yes	No	No
Country fixed effects	Yes	No	Yes	No	No
Country-approval-year fixed effects	No	Yes	No	Yes	Yes
Sector fixed effects	Yes	Yes	No	Yes	Yes
Sector-approval-year fixed effects	No	No	Yes	No	No
Approval-year fixed effects	Yes	No	No	No	No
Observations	1651	1031	1614	1003	973
R <sup>2</sup>	0.203	0.484	0.334	0.471	0.439

Note: Standard errors in parentheses; errors clustered at the evaluation-financial-year; project-level controls: DPF dummy, project amount (log), government change throughout the project and co-TTL dummy; country-level controls: corruption, ideal point (based on votes the US state department designates important) difference between recipient and US, GDP (log), FDI and democracy; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

The models mimic the specification choices explained above. Model 6 includes country, sector and approval-year fixed effects. Model 7 and 8 are estimated using country-approval-year and sector-approval-year fixed effects, respectively. Model 9 and 10 focus on government and agency performance.

While there is some evidence that staff leniency could matter, the results are not robust across the models presented. While the coefficients for leniency point in the expected direction, the coefficient is significant at conventional thresholds ( $p < 0.1$  or  $p < 0.05$ ) only in models 6 and 8. Therefore, the evidence for the importance of leniency is relatively weak. This could be for two reasons. First, leniency might not matter substantially for recipient performance. Research on the World Bank has long pointed out that staff have motivations to disburse loans due to a culture that rewards high-disbursing projects (Wapenhans 1992; Weaver 2008). Second, our measure of leniency could be imprecise. Past disbursement decisions reflect not just the individual preferences of TTLs and recipient performance but could also be influenced by other factors, like interference by powerful member states (Kilby 2013b).

The third hypothesis is evaluated through the models presented in Table 5. Again, we display the same five models, with varying fixed effects and focusing on recipient government and implementing agency performance, respectively. We present substantial evidence indicating the importance of TTLs country experience for facilitating recipient performance. The coefficient is either statistically significant or marginally significant at conventional thresholds in all five models.

So far, we focused on identifying whether staff-level factors shape recipient performance. Now, we aim to understand better whether the identified associations are meaningful. One way of gauging the relative importance of staff factors is to compare their extent with country-level explanations. We compare the two variables with relevant variables on the country-level in Table 6. To do so, we estimate two sets of regressions for each independent

**Table 5** Results of OLS models regressing recipient performance rankings on experience

	(11)	(12)	(13)	(14)	(15)
	Recipient	Recipient	Recipient	Government	Agency
Staff experience	0.0886*	0.1238***	0.0930**	0.1111**	0.0933**
	(0.0440)	(0.0396)	(0.0418)	(0.0449)	(0.0441)
Project-level controls	Yes	Yes	Yes	Yes	Yes
Country-level controls	Yes	No	Yes	No	No
Country fixed effects	Yes	No	Yes	No	No
Country-approval-year fixed effects	No	Yes	No	Yes	Yes
Sector fixed effects	Yes	Yes	No	Yes	Yes
Sector-approval-year fixed effects	No	No	Yes	No	No
Approval-year fixed effects	Yes	No	No	No	No
Observations	3134	2525	3109	2482	2440
$R^2$	0.163	0.438	0.252	0.452	0.421

Note: Standard errors in parentheses; errors clustered at the evaluation-financial-year; project-level controls: DPF dummy, project amount (log), government change throughout the project and co-TTL dummy; country-level controls: corruption, ideal point (based on votes the US state department designates important) difference between recipient and US, GDP (log), FDI and democracy; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

**Table 6** Comparing the explanatory potential of staff-level with relevant country-level factors (beta coefficients)

	(16)	(17)	(18)	(19)	(20)	(21)
Staff supervision	0.093*** (0.0360)	0.070* (0.0424)				
Staff leniency			-0.087*** (0.3494)	-0.057* (0.3632)		
Staff experience					0.040** (0.0410)	0.037* (0.0440)
GDP (log)	0.117*** (0.0109)	-0.139 (0.2038)	0.093*** (0.0113)	-0.280 (0.2124)	0.098*** (0.0113)	-0.642* (0.1825)
Democracy	-0.138*** (0.1544)	-0.099 (0.4522)	-0.146*** (0.1476)	-0.129 (0.4431)	-0.141*** (0.1189)	-0.136** (0.2963)
Corruption	-0.118*** (0.1536)	-0.123 (0.5597)	-0.110*** (0.1536)	-0.113 (0.5642)	-0.159*** (0.1270)	-0.158* (0.3866)
UNGA voting US (distance)	-0.044 (0.0417)	0.130* (0.0929)	-0.054 (0.0480)	0.158* (0.1092)	-0.035 (0.0342)	0.084 (0.0712)
FDI	0.005 (0.0028)	-0.070** (0.0044)	0.012 (0.0028)	-0.056* (0.0046)	0.016 (0.0026)	-0.046* (0.0043)
Project-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	No	Yes	No	Yes	No	Yes
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Approval-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1771	1766	1657	1651	3138	3134
R <sup>2</sup>	0.089	0.199	0.094	0.203	0.083	0.163

Note: Standard errors in parentheses; errors clustered at the evaluation-financial-year; project-level controls: DPF dummy, project amount (log), government change throughout the project and co-TTL dummy; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

variable. The first compares country and staff variables excluding country-fixed effects to contextualize the overall magnitude of the variables (Models 16, 18 and 20). The second employs country-fixed effects to focus on within-country differences (Models 17, 19 and 21). Overall, the models show that the coefficients of staff variables are of similar relevance as important country-level variables discussed in the World Bank literature. The coefficient for supervision is only slightly smaller than coefficients for corruption in Model 16, and it is about half as large as UNGA voting similarity with the United States in Model 17. The coefficient for leniency is similar to GDP (log) in strength in Model 18. However, leniency is substantially smaller than all country-level variables when employing country-fixed effects in Model 19. The same is apparent for country experience, where the coefficient is moderate compared to the main country-variables of interest. Overall, the staff-level variables are similar in magnitude as FDI inflows, the central explanatory variable in extant analyses of recipient performance (Girod and Tobin 2016). Therefore, we can say that the three staff-level variables show a sizeable association with recipient performance, although not a stronger one than the central country-level variables discussed in the World Bank literature, such as UNGA voting and corruption in recipient countries.

**Table 7** Results of OLS models regressing the staff variables on sub-samples focusing on IPF and DPF operations

	(22)	(23)	(24)	(25)	(26)	(27)
	IPF	DPF	IPF	DPF	IPF	DPF
Staff supervision	0.0827*	-0.0095				
	(0.0447)	(0.0924)				
Staff leniency			-0.4571	-0.1274		
			(0.4523)	(0.6382)		
Staff experience					0.1108**	-0.0022
					(0.0472)	(0.0707)
Project-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Country-level controls	Yes	Yes	Yes	Yes	Yes	Yes
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Sector fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Approval-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1400	341	1283	343	2569	551
R <sup>2</sup>	0.217	0.554	0.226	0.561	0.169	0.461

Note: Standard errors in parentheses; errors clustered at the evaluation-financial-year; project-level controls: DPF dummy, project amount (log), government change throughout the project and co-TTL dummy; country-level controls: corruption, ideal point (based on votes the US state department designates important) difference between recipient and US, GDP (log), FDI and democracy; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

The final part of our analysis focuses on the differences between the two major types of loans the World Bank administers. As briefly discussed above, there are central variations in the administrative procedure of different World Bank projects. To understand the scope of our arguments regarding the ability of staff to foster recipient performance, we now focus on some of these differences. We highlighted that policy-based lending, DPF, necessitates a substantially different set of actions than investment loans, IPF. Since abolishing Structural-Adjustment-Loans in 2004, the World Bank has mainly relied on DPLs. DPLs rely considerably more on so-called prior action conditions that need to be fulfilled before the project starts. After compliance with these conditions, DPLs are paid out as budget support. IPFs, on the other hand, focus on specific projects that are implemented over a more extended period of time. The World Bank expects recipients to adhere to a set of behaviors specified in the aid agreement and to perform satisfactorily. While we controlled for these differences through a dummy variable in the presented models, we also want to account for them more explicitly. In Table 7, we display results for both kinds of World Bank projects: DPFs<sup>15</sup> and IPFs. Models 19–24 replicate Models 1, 6 and 10 but focus on sub-samples in the loan type of projects. We choose the less demanding specification for these models because we split

<sup>15</sup> Of the 341 to 551 projects in the DPF sub-samples, 50 to 100 are SALs (depending on the availability of staff data). Since they do not rely as much on prior action conditionality as the newer DPFs, they would warrant additional analysis. The findings focusing only on SALs are substantially similar to the findings regarding DPF in general. We refrain from reporting these results at length, because the small sample size potentially jeopardizes robust inference.

the sample and employing country-approval-year or sector-approval-year fixed effects leads to the exclusion of a substantial number of projects that do not vary within these groups.

Table 7 shows stark differences between loan types. In the case of DPF, staff variables fail to attain statistical significance at conventional thresholds. Two interpretations appear reasonable. First, it could be that the longer time frame and the greater need for supervision make IPFs more prone to the influence of staff than DPF. Satisfactory DPF performance, in the eyes of the World Bank, includes implementing a set of prior action conditions before the project even starts. Therefore, there is much less scope for staff influence in the implementation of these projects. This finding is supported by the much larger variance the fixed effects explain in the DPF than in the IPF models. Structural factors in the design stage seem to matter more for DPF than for IPF. The second explanation is that the substantially smaller number of DPF in the sample leads to less precision when estimating the association. However, even when running bivariate OLS regressions without any control and a substantially larger sample, the coefficients remain statistically insignificant and small despite the exclusion of country-approval-year fixed effects. This gives some indication that the different findings for IPF and DPF are substantive rather than due to sample size.

## 6 Robustness checks

The robustness checks aim to rectify three possible concerns with our results. These concerns are biased ratings, selection into projects, and omitted variable bias. We discuss each of them in turn.

### 6.1 Biased ratings

First, we address the potential that the staff might be able to bias ratings. Several authors have discussed concerns with bias in the IEG ratings in particular (Kilby and Michaelowa 2019; Malik and Stone 2018) and IO evaluations in general (Eckhard and Jankauskas 2019). Others have argued that it is relatively unlikely that World Bank staff would bias the ratings because there are few incentives to do so. Staff are primarily evaluated based on running larger projects and maximizing disbursement rather than on producing well-evaluated projects (Dreher et al. 2013; Wapenhans 1992; Weaver 2008). Nevertheless, we try to minimize the possibility that bias in ratings rather than actual differences in compliance drive the correlations we discussed.

There are two potential biases: halo effect and gamed ratings. First, some authors working with IEG data have pointed to a “halo effect” where evaluators mostly care about the outcome rating and then determine the recipient and bank performance ratings accordingly (Kilby 2013a). This would imply that recipient performance ratings do not include much new information beside what is already included in outcome ratings. Therefore, we estimate models controlling for IEG outcome ratings (Appendix Table A3). We find a significant association for leniency and country experience, even when holding outcome ratings constant. The coefficient of supervision decreases substantially and fails to attain statistical significance at any conventional threshold. The robustness check implies that supervision does not explain differences between

outcome and recipient performance ratings. In light of the findings of Denizer et al. (2013) and Bulman et al. (2017), we are cautiously optimistic that this is due to the fact that “TTL quality” is also a key predictor of outcome ratings. In other words, while country experience and leniency might specifically work through influencing recipients, supervisory capacity seems to apply to other aspects of the project too.

Second, to address the possibility that staff can game ratings, we utilize the different ratings produced by IEG. It is much more difficult to bias the PPAR ratings than the ICR ratings because PPARs are done independently of the project team. Therefore, we re-estimate the models using only these ratings (Table A6). We lose a large number of observations in our sample. The directions of the coefficients are similar to the ones in the main models. Supervision and leniency are significant when excluding country fixed effects. When employing country fixed effects, none of the three variables stays significant. When we analyze the same projects ICRs, the coefficients are similarly insignificant for experience and supervision. Therefore, the finding for those two variables is likely due to sample restrictions, rather than gamed ratings. In addition, we utilize that IEG occasionally changes the rating outcome from ICRs to PPARs (Table A5). If ratings were gamed more by the staff members we focus on, a correlation between staff-level variables and negative rating changes should be observable. However, we do not find a statistically significant correlation between any of the three staff variables and rating changes. Furthermore, IEG rates the quality of the ICRs. If gamed ratings are related to the variations in staff we observe, then that could imply a lower quality of ICRs for those staff members. Therefore, we regress the ICR quality rating on the staff variables (Table A4). The approach we take is valid if one believes that IEG is able to detect gamed rating at least in some cases. Furthermore, the robustness checks we use regarding the “halo effect” also have implications for the gamed-ratings argument. If staff gamed ratings, they would surely impact the outcome rating. There is little reason to bias the recipient performance rating, while not doing so for the outcome rating. The fact that we get robust results for including the outcome ratings further increases our confidence that the results are not driven by gamed ratings.

## 6.2 Selection into projects

Another possible concern is that staff might be able to pick specific projects that are more promising. Recent research indicates that if anything, high-performing staff are allocated to more difficult contexts (Limodio 2021). Nevertheless, this is an option that warrants careful examination because we would face a selection bias. That would mean that higher compliance rates are a feature of less ambition rather than actual higher compliance. We try to account for the ambition of the project. Malik and Stone (2018) code the number of objectives each project has. While this measure does not cover substantial differences in the ambition of individual objectives, we assume that projects with more objectives are also more ambitious. This assumption could be questioned. Therefore, we employ a further indicator on the overall project costs according to the World Bank project database. Neither of our staff variables predicts project objectives. However, we do find that staff with country experience tends to run projects with lower overall costs (Table A8).

### 6.3 Omitted variables and alternative measurements

To address concerns with the calculation of the staff variables, we implement several alternative ways to calculate the staff variables in the appendix. First, we employ robustness checks to normalize the variables by active projects and use alternate data sources to construct the supervision indicators (Table A9). Second, we test for the robustness of leniency to using alternative indicators. The checks concentrate on normalization by active projects and the possibility that false equivalency was introduced through using recipient performance ratings as a denominator (Table A10). Third, we created several other versions of the country experience variable. They are used to account for criticisms regarding the aggregation of experience when co-TTLs are in charge of the project and doubts regarding the way in which experience increases networks (Table A11).

Additionally, we employ a range of different specification choices. We re-estimate the models using non-linear models. Those include a conditional logit with a binary dependent variable that codes values 1–3 as unsatisfactory and 4–6 as satisfactory. Furthermore, we use an alternative dependent variable that codes only 5 and 6 as satisfactory to account for the overall very positive ratings (Table A12). In addition, we estimate an ordered logit because the original ratings are given on an ordinal scale (Table A13). Next to these alternative model choices, we alter the original specification of the models in several ways. Modifications include varying the fixed effects used in the models to account for relevant years beyond the approval-year. We use effectiveness-year fixed effects and evaluation-year fixed effects instead of approval-year fixed effects (Table A16). Additionally, we cluster standard errors at the country level instead of the evaluation year (Table A15).

Further auxiliary control variables are employed in robustness checks. There is a trade-off in controlling for project-level factors because the staff influence most project-level variables. It could be these very factors that staff can use to increase recipient performance. On the one hand, we do not want to bias our results by accounting for project-level factors that staff use to achieve a higher recipient performance. This is the reason we are cautious in including project-level controls in the main models of this article. Nevertheless, we control for the duration of the project, sub-sector fixed effects, a count of the number of co-TTLs, and whether there was an election throughout the project (Table A14). Finally, we estimate models, including the different TTL variables together. While they are theoretically distinct, there are empirical overlaps of the different TTL competencies. Therefore, we hold the respective TTL variables constant and estimate their conditional association with recipient performance (Table A17).

The robustness checks support the conclusions drawn in the main body of the article. Supervision is statistically significant or marginally significant in most alternative specifications used. However, it fails to attain statistical significance, when controlling for IEG outcome ratings (Model 28), in some of the models focusing on the sub-sample of projects where PPARs available (Models 41 and 42) as well as when incorporating some additional project-level controls (Models 80 and 82). The evidence for leniency is much more unsteady. Leniency is not significant in nearly all models employing country-approval-year fixed effects throughout, when using any alternative indicator (Models 65–67), and when clustering standard errors at the country-level. Experience is statistically significant or marginally significant in all models used throughout the appendix, except when using the smaller sample where PPARs are available (Models 46–48). Finally, when

including all three TTL variables in Table A16, we find that supervision and leniency fail to attain significant results. This is likely because, while leniency and supervision are theoretically distinct, they overlap. TTLs that supervise more diligently might also base disbursements more on recipient performance (see appendix Table A17). Nevertheless, when holding the two other variables constant, country experience consistently attains statistically significant results (Models 99–101). Overall, we find considerable support that supervision and experience shape recipient performance in World Bank projects, both throughout the main specifications and the robustness checks. Furthermore, we find some indications that leniency might play a role.

## 7 Conclusion

The article started from the observation that there is substantial variation left unexplored in the empirical literature on recipient performance in World Bank projects. We built on research that has shown that these variations in IO service provision can matter (Smets et al. 2013) but shifted the focus to variations between international bureaucrats in their ability to shape recipient performance positively.

The findings of this article demonstrate that variation between the international bureaucrats tasked with supervision is highly relevant. In the case of the World Bank, variations between staff appear to shape recipient performance nearly as strongly as variations between countries. Staff seem to be able to influence recipient performance in a wide variety of country contexts. We explored variations between staff members by extending three perspectives on reasons for member states' adherence to IO demands to this new group of actors. Evidence from a battery of statistical tests implies three major takeaways: first, high performing staff members can foster recipient performance. Second, staff members who have better country knowledge and networks can positively contribute to recipient performance. Third, there is some evidence that leniency of staff may play a role, but it is much less robust.

One important nuance in the results is that the importance of the different staff variables varies by administrative arrangement (IPF or DPF) and by the recipient actor that the World Bank works with (government or implementing agency). We show that staff seem to matter in shaping recipient performance for IPF, which made up around US\$ 12 billion or 52% of IBRD and IDA commitments in 2020. However, the results also illustrate that the same cannot be said about DPF (US\$ 8.3 billion or 34% of World Bank operations in 2020).<sup>16</sup> Staff do not seem to be able to make a difference in persuading or aiding recipients to adhere to prior action conditionality included in DPF, that often need considerable political will. This finding could imply that World Bank staff make a difference in cases where they interact with administrative counterparts but not in cases where high-level politics shapes recipient performance. More research could aim at further unpacking these differences.

---

<sup>16</sup> The amounts are calculated on the commitments made by IDA and IBRD in 2020 according to the World Bank project and operations database as of November 2020. The remaining 14% were program for results. These loans have substantially increased in recent years but there are not enough evaluated loans to analyse them in this article.



The results presented in this article contribute to two debates in IO research. First, although the empirical section of the article focused on the case of World Bank assistance agreements, we think that our findings are potentially applicable to a larger universe of aid agencies. Recent research has found that findings generated on the case of the World Bank seem to apply to a larger number of aid agencies, both bilateral and multilateral (Briggs 2020). Furthermore, analysis of the impact of TTLs on projects of the Asian Development Bank yields similar results as those from the World Bank (Bulman et al. 2017). More research is needed to unpack how variations between international bureaucrats shape the implementation of development projects.

Second, our findings reinforce calls for extending debates on adherence to and compliance with international agreements beyond state-centric accounts to include the variety of actors that are crucial for compliance processes (Dai 2005; Mitchell and Hensel 2007). In their adherence to World Bank assistance agreements, recipient governments and their implementing agencies seem to be influenced by international bureaucrats. This implication seems the more important as the positive contributions to compliance seem to occur irrespective of the country context. Therefore, the compliance literature could benefit from engaging more with debates on the varying influence of international bureaucrats (Bayerlein et al. 2020; Chwieroth 2013; Eckhard and Ege 2016).

Finally, the findings imply that the World Bank and other development agencies could benefit from re-thinking their organizational incentive structure for staff. It seems to matter who is supervising the project. More targeted staffing policies could, therefore, help to foster recipient performance. Better performing staff and staff with more country experience seem to be able to shape recipient performance. Recent research shows that World Bank staff think that their career advances are primarily dependent on moving money (Briggs 2019a). The findings of this article reinforce calls for changing the incentives of staff to focus more on skills and networks of staff, rather than getting large projects approved and disbursed. Reforming staff incentive structure to reward those that contribute to the organizational mandate could prove beneficial for the World Bank.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11558-021-09414-4>.

**Acknowledgements** We thank Louisa Bayerlein, Martin Binder, Ben Christian, Thomas Dörfler, Timon Forster, Vytautas Jankauskas, Mathies Kempken, Christopher Kilby, Christoph Knill, Nina Reiners, Ann-Kathrin Rothermel, Alexandros Tokhi, Matthew Winters and three anonymous reviewers for constructive comments. Previous versions of this article were presented at the APSA Annual Meeting 2020 and the ECPR General Conference 2020 and we are grateful to participants for their comments and suggestions. All remaining errors are our own. Clara Bünger and Mathias Tasser provided excellent research assistance for this article.

**Funding** Funding by the German Research Foundation (DFG) as part of the Research Unit International Public Administration (IPA) with grant number FOR#1745 TP LI 1947/4–1 is gratefully acknowledged.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aguilar, M. A., Gill, J. B. S., & Pino, L. (2010). Preventing fraud and corruption in World Bank projects. A Guide for Staff. Washington, DC: World Bank. <http://citeseerx.ist.psu.edu/viewdoc/download?https://doi.org/10.1.1.457.371&rep=rep1&type=pdf>.
- Andersen, T. B., Hansen, H., & Markussen, T. (2006). US politics and World Bank IDA-lending. *The Journal of Development Studies*, 42(5), 772–794. <https://doi.org/10.1080/00220380600741946>.
- Arpac, O., & Bird, G. (2009). Turkey and the IMF: A case study in the political economy of policy implementation. *Review of International Organizations*, 4(2), 135–157. <https://doi.org/10.1007/s11558-008-9051-4>.
- Babb, S. (2009). *Behind the development banks. Washington politics, world poverty, and the wealth of nations*. Chicago and London: University of Chicago Press.
- Bailey, M. A., Strezhnev, A., & Voeten, E. (2017). Estimating dynamic state preferences from United Nations voting data. *Journal of Conflict Resolution*, 61(2), 430–456. <https://doi.org/10.1177/0022002715595700>.
- Barnett, M. N., & Finnemore, M. (2004). *Rules for the world. International organizations in global politics*. Cornell: Cornell University Press.
- Bayerlein, L., Knill, C., & Steinebach, Y. (2020). *A matter of style. Organizational agency in global public policy*. Cambridge: Cambridge University Press.
- Bazbauers, A. (2019). World Bank technical assistance: Locating and cultivating sympathetic interlocutors. *Development Policy Review*, 37(5), 638–656. <https://doi.org/10.1111/dpr.12377>.
- Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106(2), 265–289. <https://doi.org/10.1037/0033-2909.106.2.265>.
- Börzel, T. A. (2020). *Why noncompliance. The politics of law in the European Union*. Ithaca: Cornell University Press.
- Börzel, T. A., Hofmann, T., Panke, D., & Sprungk, C. (2010). Obstinate and inefficient: Why member states do not comply with European law. *Comparative Political Studies*, 43(11), 1363–1390. <https://doi.org/10.1177/0010414010376910>.
- Briggs, R. C. (2019a). Why does aid not target the poorest? *Aiddata Working Paper 85*. Williamsburg, VA: Aiddata. [http://docs.aiddata.org/ad4/pdfs/WPS85\\_Why\\_does\\_aid\\_not\\_target\\_the\\_poorest.pdf](http://docs.aiddata.org/ad4/pdfs/WPS85_Why_does_aid_not_target_the_poorest.pdf).
- Briggs, R. C. (2020). Results from single-donor analyses of project aid success seem to generalize pretty well across donors. *Review of International Organizations*, 15, 947–963. <https://doi.org/10.1007/s11558-019-09365-x>.
- Broome, A., & Seabrooke, L. (2015). Shaping policy curves. Cognitive authority in transnational capacity building. *Public Administration*, 93(4), 956–972. <https://doi.org/10.1111/padm.12179>.
- Bueno De Mesquita, B., & Smith, A. (2009). A political economy of aid. *International Organization*, 63(2), 309–340. <https://doi.org/10.1017/S0020818309090109>.
- Bulman, D., Kolkma, W., & Kraay, A. (2017). Good countries or good projects? Comparing macro and micro correlates of World Bank and Asian Development Bank project performance. *Review of International Organizations*, 12, 335–363. <https://doi.org/10.1007/s11558-016-9256-x>.
- Chayes, A., & Chayes, A. H. (1993). On compliance. *International Organization*, 47(2), 175–205. <https://doi.org/10.1017/S0020818300027910>.
- Checkel, J. T. (1997). International norms and domestic politics: Bridging the rationalist-constructivist divide. *European Journal of International Relations*, 3(4), 473–495. <https://doi.org/10.1177/1354066197003004003>.
- Chwieroth, J. M. (2013). “The silent revolution:” how the staff exercise informal governance over IMF lending. *Review of International Organizations*, 8(2), 265–290. <https://doi.org/10.1007/s11558-012-9154-9>.
- Clark, R., & Dolan, L. R. (2021). Pleasing the principal: U.S. influence in World Bank policymaking. *American Journal of Political Science*, 65(1), 36–51. <https://doi.org/10.1111/ajps.12531>.

- Copelovitch, M. S. (2010). Master or servant? Agency slack and the politics of IMF lending. *International Studies Quarterly*, 54, 49–77. <https://doi.org/10.1111/j.1468-2478.2009.00577.x>.
- Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., Altman, D., et al. (2019). V-Dem [country-year] dataset v9. *Varieties of Democracy (V-Dem) Project*. <https://www.v-dem.net/en/data/data-version-9/>. Accessed 1 January 2020.
- Cormier, B. (2016). Empowered borrowers? Tracking the World Bank's program-for-results. *Third World Quarterly*, 37(2), 209–226. <https://doi.org/10.1080/01436597.2015.1108828>.
- Dai, X. (2005). Why comply? The domestic constituency mechanism. *International Organization*, 59(2), 363–398. <https://doi.org/10.1017/S0020818305050125>.
- Denizer, C., Kaufmann, D., & Kraay, A. (2013). Good countries or good projects? Macro and micro correlates of World Bank project performance. *Journal of Development Economics*, 105, 288–302. <https://doi.org/10.1016/j.jdeveco.2013.06.003>.
- Dietrich, S. (2013). Bypass or engage? Explaining donor delivery tactics in foreign aid allocation. *International Studies Quarterly*, 57(4), 698–712. <https://doi.org/10.1111/isqu.12041>.
- Dollar, D., & Levin, V. (2005). Sowing and reaping: Institutional quality and project outcomes in developing countries. *Policy Research Working Paper 3524*. Washington, DC: World Bank. <https://openknowledge.worldbank.org/bitstream/handle/10986/8885/wps3524.pdf?sequence=1&isAllowed=y>.
- Dollar, D., & Svensson, J. (2000). What explains the success or failure of structural adjustment programmes? *The Economic Journal*, 110(466), 894–917. <https://doi.org/10.1111/1468-0297.00569>.
- Downs, G. W. (1998). Enforcement and the evolution of cooperation. *Michigan Journal of International Law*, 19(2), 319–344. <https://repository.law.umich.edu/mjil/vol19/iss2/2>.
- Dreher, A., Klasen, S., Vreeland, J. R., & Werker, E. (2013). The costs of favoritism: Is politically driven aid less effective? *Economic Development and Cultural Change*, 62(1), 157–191. <https://doi.org/10.1086/671711>.
- Easterly, W. (2014). *The tyranny of experts: Economists, dictators, and the forgotten rights of the poor*. New York: Basic Books.
- Eckhard, S. (2020). Bridging the citizen gap: Bureaucratic representation and knowledge linkage in (international) public administration. *Governance*, 1–20. <https://doi.org/10.1111/gove.12494>.
- Eckhard, S., & Ege, J. (2016). International bureaucracies and their influence on policy-making: A review of empirical evidence. *Journal of European Public Policy*, 23(7), 960–978. <https://doi.org/10.1080/13501763.2016.1162837>.
- Eckhard, S., & Jankauskas, V. (2019). The politics of evaluation in international organizations: A comparative study of stakeholder influence potential. *Evaluation*, 25(1), 62–79. <https://doi.org/10.1177/1356389018803967>.
- Fearon, J. D. (1998). Bargaining, enforcement, and international cooperation. *International Organization*, 52(2), 269–305. <https://doi.org/10.1162/002081898753162820>.
- Girod, D. M., & Tobin, J. L. (2016). Take the money and run: The determinants of compliance with aid agreements. *International Organization*, 70, 209–239. <https://doi.org/10.1017/S0020818315000326>.
- Guimaraes, P., & Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *The Stata Journal*, 10(4), 628–649.
- Hawkins, D. G., Lake, D. A., Nielson, D. L., & Tierney, M. J. (2006). *Delegation and Agency in International Organizations*. Cambridge: Cambridge University Press.
- Heinzel, M., Richter, J., Busch, P., Feil, H., Herold, J., & Liese, A. (2020). Birds of a feather? The determinants of impartiality perceptions of the IMF and the World Bank. *Review of International Political Economy*, 1–25. <https://doi.org/10.1080/09692290.2020.1749711>.
- Honig, D. (2019). When reporting undermines performance: The costs of politically constrained organizational autonomy in foreign aid implementation. *International Organization*, 73, 171–201. <https://doi.org/10.1017/S002081831800036X>.
- Honig, D. (2020). Information, power, & location: World Bank staff decentralization and aid project success. *Governance*, 33(4), 749–769. <https://doi.org/10.1111/gove.12493>.
- Hurd, I. (1999). Legitimacy and Authority in International Politics. *International Organization*, 53(2), 379–408. <https://doi.org/10.1162/002081899550913>.
- Ika, L. A., Diallo, A., & Thuillier, D. (2012). Critical success factors for World Bank projects: An empirical investigation. *International Journal of Project Management*, 30(1), 105–116. <https://doi.org/10.1016/j.ijproman.2011.03.005>.
- Keefe, P., & Knack, S. (2008). Social capital, social norms and the new institutional economics. In C. Menard & M. M. Shirley (Eds.), *Handbook of new institutional economics* (Vol. 57, pp. 701–725). Berlin and Heidelberg: Springer VS. <https://doi.org/10.1080/03585520903122574>.

- Kilby, C. (2001). World Bank-borrower relations and project supervision. *Canadian Journal of Development Studies*, 22(1), 191–218. <https://doi.org/10.1080/02255189.2001.9668807>.
- Kilby, C. (2013a). The political economy of project preparation: An empirical analysis of World Bank projects. *Journal of Development Economics*, 105, 211–225. <https://doi.org/10.1016/j.jdeveco.2013.07.011>.
- Kilby, C. (2013b). An empirical assessment of informal influence in the World Bank. *Economic Development and Cultural Change*, 61(2), 431–464. <https://doi.org/10.1086/668278>.
- Kilby, C. (2015). Assessing the impact of World Bank preparation on project outcomes. *Journal of Development Economics*, 115, 111–123. <https://doi.org/10.1016/j.jdeveco.2015.02.005>.
- Kilby, C., & Michaelowa, K. (2019). What influences World Bank project evaluations? In C. R. Williamson & N. Dutta (Eds.), *Lessons on foreign aid and economic development. Micro and macro perspectives* (pp. 109–150). London: Palgrave Macmillan.
- Knack, S. (2013). Aid and donor trust in recipient country systems. *Journal of Development Economics*, 101(1), 316–329. <https://doi.org/10.1016/j.jdeveco.2012.12.002>.
- Knill, C., Bayerlein, L., Enkler, J., & Grohs, S. (2019). Bureaucratic influence and administrative styles in international organizations. *Review of International Organizations*, 14(1), 83–106. <https://doi.org/10.1007/s11558-018-9303-x>.
- Koliev, F., Sommerer, T., & Tallberg, J. (2020). Compliance without coercion: Effects of reporting on international labor rights. *Journal of Peace Research*. <https://doi.org/10.1177/0022343320910254,0022343320910254>.
- Kwan, L. Y. Y., Yap, S., & Chiu, C. Y. (2015). Mere exposure affects perceived descriptive norms: Implications for personal preferences and trust. *Organizational Behavior and Human Decision Processes*, 129, 48–58. <https://doi.org/10.1016/j.obhdp.2014.12.002>.
- Lannon, J., & Walsh, J. N. (2020). Project facilitation as an active response to tensions in international development programmes. *International Journal of Project Management*, <https://doi.org/10.1016/j.ijproman.2020.06.002>.
- Limodio, N. (2021). Bureaucrats allocation in the public sector: Evidence from the World Bank. *The Economic Journal*. <https://doi.org/10.1093/ej/ueab008>.
- Lipsky, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York: Russell Sage Foundation.
- Malik, R., & Stone, R. W. (2018). Corporate influence in World Bank lending. *The Journal of Politics*, 80(1), 103–118.
- Marchesi, S., & Masi, T. (2020). Delegation of implementation in project aid. *Review of International Organizations*. <https://doi.org/10.1007/s11558-020-09396-9>.
- May, P. J., & Winter, S. C. (2009). Politicians, managers, and street-level bureaucrats: Influences on policy implementation. *Journal of Public Administration Research and Theory*, 19(3), 453–476. <https://doi.org/10.1093/jopart/mum030>.
- Mitchell, S. M. L., & Hensel, P. R. (2007). International institutions and compliance with agreements. *American Journal of Political Science*, 51(4), 721–737. <https://doi.org/10.1111/j.1540-5907.2007.00277.x>.
- Raustiala, K., & Slaughter, A.-M. (2002). International law, international relations and compliance. In W. Carlsnaes, T. Risse, & B. A. Simmons (Eds.), *Handbook of international relations*. Thousand Oaks: SAGE Publications, 538–558.
- Scartascini, C., Cruz, C., & Keefer, P. (2017). *The Database of Political Institutions 2017*. Washington, DC: Inter-American Development Bank. <https://doi.org/10.18235/0001027>.
- Simmons, B. A. (1998). Compliance with international agreements. *Annual Review of Political Science*, 1(1), 75–93. <https://doi.org/10.1146/annurev.polisci.1.1.75>.
- Simmons, B. A. (2009). *Mobilizing for human rights: International law in domestic politics*. Cambridge: Cambridge University Press.
- Simmons, B. A. (2013). International law. In W. Carlsnaes, T. Risse, & B. A. Simmons (Eds.), *Handbook of international relations (2nd ed.)*. Thousand Oaks: SAGE Publications, 352–378.
- Smets, L., Knack, S., & Molenaers, N. (2013). Political ideology, quality at entry and the success of economic reform programs. *Review of International Organizations*, 8, 447–476. <https://doi.org/10.1007/s11558-013-9164-2>.
- Stone, R. W. (2002). *Lending credibility: The International Monetary Fund and the post-communist transition*. Princeton: Princeton University Press.
- Stone, R. W. (2013). Informal governance in international organizations: Introduction to the special issue. *Review of International Organizations*, 8, 121–136. <https://doi.org/10.1007/s11558-013-9168-y>.
- Tallberg, J. (2002). Paths to compliance: Enforcement, management, and the European Union. *International Organization*, 56(3), 609–643. <https://doi.org/10.1162/002081802760199908>.

- Vreeland, J. R. (2006). IMF program compliance: Aggregate index versus policy specific research strategies. *Review of International Organizations*, 1(4), 359–378. <https://doi.org/10.1007/s11558-006-0161-6>.
- Vreeland, J. R. (2019). Corrupting international organizations. *Annual Review of Political Science*, 22(1), 205–222. <https://doi.org/10.1146/annurev-polisci-050317-071031>.
- Wade, R. H. (2002). US hegemony and the World Bank: The fight over people and ideas. *Review of International Political Economy*, 9(2), 215–243. <https://doi.org/10.1080/0969229011012609>.
- Wane, W. (2004). The quality of foreign aid: Country selectivity or donors incentives? *World Bank Working Paper* 3325. Washington, DC: World Bank. <https://openknowledge.worldbank.org/bitstream/handle/10986/14001/wps3325.pdf?sequence=1&isAllowed=y>.
- Wapenhans, W, A. (1992). Effective implementation: Key to development impact. World Bank: Washington, DC. <http://documents1.worldbank.org/curated/en/596521467145356248/pdf/multi0page.pdf>.
- Weaver, C. (2008). *Hypocrisy trap. The World Bank and the poverty of reform*. Princeton and Oxford: Princeton University Press.
- Winters, M. S. (2014). Targeting, accountability and capture in development projects. *International Studies Quarterly*, 58(2), 393–404. <https://doi.org/10.1111/isqu.12075>.
- Winters, M. S. (2019). Too many cooks in the kitchen? The division of financing in World Bank projects and project performance. *Politics and Governance*, 7(2), 117–126. <https://doi.org/10.17645/pag.v7i2.1826>.
- Wolfensohn, J. D. (1996). People and development. Washington, DC. World Bank. <http://documents1.worldbank.org/curated/en/135801467993234363/pdf/People-and-development-annual-meetings-address-by-James-D-Wolfensohn-President.pdf>
- Woods, N. (2006). *The globalizers. The IMF, the World Bank, and their borrowers*. Ithaca: Cornell University Press.
- World Bank. (2013a). Task team leader toolkit : Investment project financing - the project cycle. Washington, DC: World Bank. <http://documents1.worldbank.org/curated/en/986001468010231537/pdf/951680WP0P12930BLIC00webTTL0toolkit.pdf>.
- World Bank. (2013b). *Transcript of oral history interview with Paul Cadario held on April 4, 17, 22, 30, 2013*. Washington, DC: World Bank. <https://oralhistory.worldbank.org/transcripts/transcript-oral-history-interview-paul-cadario-held-april-4-17-22-30-2013>.
- World Bank. (2015a). Learning and results in World Bank operations: Toward a new learning strategy. Washington, DC: World Bank IEG. [https://ieg.worldbankgroup.org/sites/default/files/Data/reports/chapters/LR2\\_full\\_report\\_revised.pdf](https://ieg.worldbankgroup.org/sites/default/files/Data/reports/chapters/LR2_full_report_revised.pdf).
- World Bank. (2015b). World Bank Project Performance Ratings -Codebook. Washington, DC: World Bank IEG. [https://ieg.worldbankgroup.org/sites/default/files/Data/reports/ieg-wb-project-performance-ratings-codebook\\_092015.pdf](https://ieg.worldbankgroup.org/sites/default/files/Data/reports/ieg-wb-project-performance-ratings-codebook_092015.pdf).
- World Bank. (2016). *Behind the mirror: A report on the self-evaluation systems of the World Bank Group*. Washington, DC: World Bank. [https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/behindthemirror\\_0716.pdf](https://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/behindthemirror_0716.pdf).
- World Bank. (2018). *Transcript of oral history interview with Jan Walliser held on June 13, 2018*. Washington, DC: World Bank. <https://oralhistory.worldbank.org/transcripts/transcript-oral-history-interview-jan-walliser-held-june-13-2018>
- World Bank. (2020a). IEG World Bank Project Performance Ratings. Washington, DC: World Bank IEG. <https://datacatalog.worldbank.org/dataset/ieg-world-bank-project-performance-ratings>.
- World Bank. (2020b). World Development Indicators. Washington, DC: World Bank. <http://datatopics.worldbank.org/world-development-indicators/>.
- Yelle, L. E. (1979). The learning curve: Historical review and comprehensive survey. *Decision Sciences*, 10(2), 302–328. <https://doi.org/10.1111/j.1540-5915.1979.tb00026.x>.