AIDDATA
A Research Lab at William & Mary

WORKING PAPER 40

May 2017

When Do Global Performance Assessments Influence Policy Behavior? Micro-Evidence from the 2014 Reform Efforts Survey

Bradley C. Parks
Institute for the Theory and Practice of International Relations, The College of William & Mary

Takaaki Masaki
Institute for the Theory and Practice of International Relations, The College of William & Mary

Abstract

Scholars and policymakers generally agree that global performance assessments (GPAs) can influence the policy priorities and actions of public sector decision-makers. However, there is little systematic evidence about the conditions under which GPAs — and performance assessments, more generally — instigate changes in state behavior. There is also a lack of understanding about the causal mechanisms through which GPAs and other types of performance assessments facilitate policy changes. We seek to close this evidence gap by leveraging a survey of 1,788 host government officials that provides comparative data on the agenda-setting influence and reform design influence of more than 100 government performance assessments. We argue that GPAs function as signaling devices that provide credibility assurances to foreign investors and donors. However, the net benefits of credibility signaling to these external actors must be sufficiently large and certain for policymakers in assessed countries to recalibrate their domestic reform priorities and efforts. We posit that this condition is met when the supplier of a cross-country performance assessment allows assessed governments to participate in the assessment process. Using a multilevel linear model to account for the hierarchical structure of our survey data, we find evidence that performance assessments yield greater policy influence when they make an explicit comparison of government performance across countries and allow assessed governments to participate in the assessment process. This finding is robust to a variety of tests, including country-fixed and respondent-fixed effects.

Author Information

Bradley C. Parks
Institute for the Theory and Practice of International Relations and The College of William and Mary
bcpark@wm.edu

Takaaki Masaki
Institute for the Theory and Practice of International Relations and The College of William and Mary
tmasaki@wm.edu

Acknowledgments

We thank Judith Kelley, Beth Simmons, Dan Honig, Eddy Malesky, Elise Le, Steven Bernstein, and Anton Strezhnev for providing helpful comments on previous drafts of the manuscript. We also thank NORC at the University of Chicago — in particular, Renee Hendley, Kate Hobson, David Kordus, Elise Le, Aparna Ramakrishnan, Alex Rigaux, Bhanuj Soni, and Stacy Stonich — for their survey design and data collection assistance. Additionally, we owe a debt of gratitude to the nearly 6,750 development policymakers and practitioners from 126 countries who participated in the 2014 Reform Efforts Survey. Were it not for their willingness to graciously share of their experiences and insights, this study would not have been possible. Finally, this study was made possible through generous financial support from the John Templeton Foundation and the Smith Richardson Foundation.

Contents

1. Introduction

Between 2004 and 2012, the Government of Georgia—led by President Mikheil Saakashvili—launched a series of far-reaching reforms to overhaul tax and customs administration, business and property registration, the courts, the police, and the civil service.[1] In parallel, the Georgian authorities launched a major effort to communicate their reform accomplishments to foreign investors and dislodge prevailing assumptions about the country's policy environment and its attractiveness as an investment destination (Schueth 2011; Di Puppo 2014).[2] The authorities were particularly aggressive in their use of the World Bank's Doing Business Report to broadcast the government's reform credentials to the outside world (De Waal 2011). They launched a multi-million dollar publicity campaign, purchasing print and television advertisements with The Economist, the Financial Times, CNN, CNBC, and other high-profile financial media outlets (Schueth 2011). With a slogan of "And the winner is: GEORGIA", the campaign compared Georgia's very favorable performance on the Doing Business rankings with the less favorable performance of developed countries (e.g. Germany) and others popular investment destinations (e.g. China). The European Stability Initiative (2010) notes that "[r]ising in [the Doing Business] rankings became a national policy priority, closely monitored by [President] Saakashvili."

The government was successful in executing its rebranding strategy and convincing foreign investors that they should reevaluate their assumptions about Georgia: Georgia catapulted from 137th place in 2004 to 18th place in 2008 on the "Ease of Doing Business Index" (EDBI) and surged ahead of Canada, Switzerland, Austria, and the Netherlands in the global rankings (USAID 2009). During this period of time, Standard and Poor's dubbed Georgia a "frontier market" where "potentially high rates of return

[1] Among other things, the Saakashvili administration fired 15,000 members of the notoriously corrupt police force, dramatically increased the salaries of 10,000 public servants to combat petty corruption, reduced the time and cost of property registration by 75%, cut the corporate income tax rate from 20% to 15%, streamlined customs by reducing the number of days it takes to export from 54 to 10, and slashed the number of activities requiring a business license from 909 to 159 (Bagaudinova et al. 2007; World Bank 2012; Bendukidze and Saakhashvili 2014; Light 2014).

[2] Wheatley (2005) argues that, prior to the Rose Revolution and the Saakashvili administration, Georgia met almost any reasonable definition of a "failed state." President Saakashvili was directly involved in the government's rebranding strategy and seemingly never missed an opportunity to remind the international community of Georgia's progress vis-à-vis its status as a "failed state" prior to the Rose Revolution. For example, in 2010, he penned an op-ed in *Foreign Policy* magazine entitled "Failed No Longer," and in it he noted that "[j]ust a decade ago, Georgia could not seriously speak of playing any constructive role in the world. We were, to be perfectly frank and accurate, a failed state. Major areas of our country were effectively run by warlords. The police system was corrupt beyond description, with state police extorting payments from prisoners' families and shakedowns by traffic police at almost every corner. … Against that backdrop, our Rose Revolution in 2003 was not just about waving flags and storming the Parliament. It was a decision to move our state from failed to functioning" (Saakashvili 2010).

may be reaped on investments that are associated with decreasing levels of risk" (Schueth 2011: 61).[3] Georgia also issued its first international sovereign debt offering in 2008 and the market response was overwhelmingly positive: the $500 million offering was oversubscribed by a factor of three. By 2007, FDI inflows tripled, tax revenues jumped by 300%, and economic growth soared above 10% (Mitchell 2009; IMF 2009).

Scholars of international relations (IR), international organization (IO), and international political economy (IPE) have for several decades argued that treaty ratification, IO membership, and voluntary submission to policy surveillance by multilateral institutions are the primary mechanisms through which credibility-seeking governments send signals of reassurance to external parties to win their trust. The Government of Georgia had most if not all of these credibility signaling mechanisms at its disposal, but it chose not to pursue any of these tools that states purportedly use to seek and acquire credibility. Instead, the Georgian authorities used a global performance assessment (GPA) to make their case to foreign investors. This strategic choice represents a growing trend and a broader puzzle for IR scholars: why do more and more countries seem to favor global performance assessments (GPAs) as their preferred signaling devices?

Our argument in this study is that GPAs function as signaling devices that provide credibility assurances to external actors such as foreign investors and donors, but that they are different from the traditional signaling devices and conventional forms of policy surveillance and assessment that states, IOs, and NGOs have undertaken for decades in that they involve rating and ranking the relative performance of countries. It is the cross-national and comparable nature of GPAs that enables governments in assessed countries to differentiate themselves from their peers and competitors, and thereby increase their odds of securing more external resources (e.g., aid, investment).[4]

However, we do not propose that GPAs on their own provide such large and certain signaling benefits that they directly and unconditionally influence the reform priorities and efforts of governing elites in assessed countries. We instead argue that policymakers must balance the credibility signaling benefits

[3] Georgia's rapid rise on the rankings in the *Doing Business* rankings also helped the government secure external validation from states and international organizations (Di Puppo 2014). The U.S. Government, the World Bank, and various donors lavishly praised Georgia during the Saakashvili administration, identifying it as a "model of reform" (Dinmore 2005) and "[something] we hope to see replicated in many other parts of the world" (Taylor 2004). USAID (2009) has characterized Georgia's rapid rise on the *Doing Business* indicators as evidence of "the broadest, deepest, fastest business climate reforms [undertaken by] any country in the last 50 years."

[4] As such, this study focuses on how GPAs operate by activating transnational pressures and altering market expectations (Kelley and Simmons 2016: 31).

of improving their performance on a GPA against the costs of foregone policy autonomy, and they often face considerable uncertainty about the net benefits that they will reap (within a politically relevant time horizon) if they address the performance requirements of a GPA. Given the difficulty that governments face in reliably estimating expected credibility signaling benefits relative to foregone domestic policy discretion, we argue that they favor and seek out GPAs that give them a higher level of certainty that the policy adjustments that they pursue will result in the signaling benefits that they hope to achieve. GPAs that allow for assessed governments to participate in the assessment process provide this higher level of certainty. They increase policymakers' confidence that, if they address the performance requirements of the assessment provider, they will reap the signaling benefits that they hope to achieve.

By identifying the specific conditions under which GPAs exert policy influence and illuminating the causal mechanisms through which GPAs operate, this study contributes to a broader body of research on the role of information in IR. Previous studies have shown that efforts to "name and shame" states—by collecting and disclosing information about their performance—can trigger policy adjustments by public sector decision-makers under certain conditions (Hafner-Burton 2008; Murdie and Peksen 2015). A related literature has demonstrated that various forms of policy surveillance can have material consequences in the states that are being monitored (Lebovic and Voeten 2009; Biglaiser and DeRouen Jr. 2010; Brooks et al. 2015).[5] It also now broadly accepted that states use a variety of signaling mechanisms—including credit ratings, treaty ratifications, membership in IOs, and compliance with policy conditions set by external actors—to transmit information about their credibility to other states, IOs, and market actors (Elkins et al. 2008; Dreher and Voigt 2011; Girod and Tobin 2016).[6] However, the existing literature is relatively silent on the question of whether specific attributes of credibility signaling mechanisms and entities amplify or constrain their policy influence in credibility-seeking states. Nor does the literature provide much systematic evidence about the causal processes through which credibility signaling mechanisms and entities operate in credibility-seeking states. Our study helps fill these gaps.

[5] Some argue that these material consequences motivate policy responses by states (Noland 1997; Öhler et al. 2012). Others argue that rather than making a rational calculation of whether a particular type of policy adjustment serves their own interests or the national interest, public sector decision-makers take "cognitive shortcuts" and look to their neighbors or other peers for information about the most "modern" and "appropriate" types of policies (Weyland 2005; Gray 2013).

[6] In fact, the collection and publication of information about state performance can lead to "gaming" or other forms of opportunistic behavior (Samuel 2014; Kerner et al. 2015; Buntaine et al. forthcoming).

There are six remaining sections of this paper. In Section II, we present a theory of GPA policy influence and identify its observable implications. In Section III, we describe the data and methods that we will use to evaluate the testable hypotheses that follow from our theory. In Sections IV and V, we introduce and interpret econometric and descriptive evidence from a new survey of 1,788 host government officials that provides comparative data on the agenda-setting influence and reform design influence of 103 performance assessments in 123 low-income and middle-income countries and territories. In Section VI, we identify and demonstrate the plausibility of a causal mechanism that may account for our results. Section VII concludes with a broader discussion of how our findings speak to the question of whether GPAs represent "new wine in old wineskins" (i.e. a 21st century manifestation of external policy surveillance and assessment activities that began in the second half of the 20th century) or new, under-explained, and substantively important phenomena in world politics that merit greater theoretical and empirical attention.

2. Theory and Hypotheses

A defining feature of governance in developing countries is the need to manage countervailing interests and pressures from external and domestic sources (Montinola 2010; Bueno de Mesquita and Smith 2010; Hollyer and Rosendorff 2012; Lake and Farris 2014). On one hand, leaders have incentives to preserve their domestic policy autonomy and maneuverability. Elected and unelected leaders need to be able to flexibly respond to key domestic constituencies and changing domestic circumstances if they wish to stay in power (Smith and Vreeland 2006; Morrison 2009; Bueno de Mesquita and Smith 2010; Quiroz Flores and Smith 2013). On the other hand, leaders often need access to resources controlled by external actors—e.g. foreign aid and investment, the ability to borrow from private and official sources, preferential access to the markets of important trading partners, international praise and legitimacy—in order to achieve their domestic objectives (Wright 2008; Licht 2010; Arias et al. 2014; Girod and Tobin 2016).[7] Therefore, an important baseline assumption from the existing theoretical literature is that when governing elites in developing countries contemplate the possibility of addressing the policy demands of an external actor (e.g. a GPA provider), they will weigh the benefits of unlocking external resources against the costs of relinquishing some degree of domestic

[7] This need to balance and trade-off countervailing pressures from domestic and external sources is usually present, regardless of whether leaders are seeking to achieve political objectives (e.g. generating sufficiently large rents to cement support within the winning coalition) or economic objectives (e.g. increasing economic growth).

policy discretion (Pevehouse 2002; Elkins et al. 2006; Büthe and Milner 2008; Hollyer and Rosendorff 2012).

The fact that some governments, after weighing these benefits and costs, still make efforts to secure material or reputational rewards from external actors has prompted theorization about whether, when, why and how states seek and acquire credibility (Martin 2000; Pevehouse 2002; Elkins et al. 2006; Büthe and Milner 2008). Many scholars argue that credibility considerations are particularly important when states seek to acquire foreign direct investment. Dreher and Voigt (2011: 327) explain why this might be the case: "[c]redibility will affect whether and [under] which conditions the government gets access to international and domestic capital.… [A]ctors will invest more than if the government was not credible… and [h]igher investment levels translate into additional income. This, in turn, leads to higher utility levels for both the governed and the governing because higher (aggregate) income also means increased tax revenue."[8] A related argument is that developing countries, in particular, have a need for credibility because they are generally less willing and able to provide reliable information on local investment climate conditions to foreign investors (Garriga and Phillips 2013; Lee forthcoming).

If one accepts this theoretical reasoning, a logical corollary is that developing countries need mechanisms to transmit signals to external actors about their credibility. Previous research points to a range of different mechanisms through which countries can signal their credibility, including membership in international organizations (Pevehouse 2002; Gray 2009; Fang and Owen 2011), the ratification of international treaties (Büthe and Milner 2008; Elkins and Simmons 2008), voluntary submission to policy surveillance or evaluation by a multilateral institution (Dhonte 1997; Vreeland 2003; Biglaiser and DeRouen 2010), the adoption of information disclosure commitments (Pitlik et al. 2010; Hollyer et al. 2011; Krasner 2011; David-Barrett and Okamura 2016), the appointment of technocratic leaders (Chwieroth 2007a, 2007b; Flores et al. 2015), the pursuit of de jure and de facto reform efforts (Marchesi and Thomas 1999; Andrews 2013; Buntaine et al. forthcoming), compliance with conditional aid contracts (Dreher et al. 2012; Girod and Tobin 2016), and the creation of strong public financial management systems to safeguard incoming aid flows (Knack 2013; Garriga and

[8] States, of course, also seek to send credibility signals to foreign donors (Dreher et al. 2012; Andrews 2013; Knack 2013; Buntaine et al. forthcoming). However, in allocating the money and external validation that recipient countries value, donors respond primarily to factors other than recipient performance (Alesina and Dollar 2000; Neumayer 2003; Easterly and Williamson 2011). Therefore, incentives to send signals of credibility, or good performance, should be weaker for donors than for investors.

Phillips 2013).[9]

For the specific purpose of credibility signaling to external actors, GPAs should be uniquely attractive mechanisms because they allow states to clearly differentiate their performance from the performance of other states and to both quickly and broadly transmit these signals (Kelley and Simmons 2016). The ultimate purpose of credibility signaling is "to separate the trustworthy types from the untrustworthy types" (Kydd 2000: 326). However, many of the traditional mechanisms for credibility signaling only allow the recipients of the information that is being transmitted to draw crude and slow-moving distinctions between trustworthy and untrustworthy types (Hyde 2015). Whereas the noise-to-signal ratio may be quite high when a state seeks to communicate its intentions or credentials through the ratification of an international treaty or participation in an IMF program, GPAs significantly lower this ratio by providing exact, comparable, and (in many cases) frequently updated information about the credibility of assessed states. In doing so, GPAs effectively allow for better separation of types, which is the primary purpose of credibility signaling.

The speed and precision of credibility signaling mechanisms is particularly important to policymakers in assessed countries because of their political time horizons. Governing elites have limited time, money, and political capital, and if they wish to send a costly signal of reassurance to external actors, they will seek to ensure that they can reap the rewards (i.e. more investment) of any costly policy adjustment(s) that they choose to undertake within a politically-relevant time horizon (e.g. before they leave office). This is particularly true in developing countries, where the average tenure of a finance minister is less than two years (Nelson 2009). Therefore, ceteris paribus, one would expect governments to derive larger benefits from—and therefore generally favor—credibility signaling mechanisms that are (a) more sensitive to domestic policy adjustments, and (b) less noisy ways of differentiating their intentions and credentials from those of their peers and competitors (Schimmelfennig and Sedelmeier 2004; Parks 2014; Custer et al. 2016; Kelley and Simmons 2016). GPAs should for these reasons be more attractive than the alternative mechanisms that governments can use to transmit credibility signals to external actors.

---

[9] Some of these signals are more costly and credible than others. For example, compliance with IMF policy conditionality sends a more potent and persuasive signal than IMF program participation (Mody and Saravia 2003); membership in an international organization that promotes human rights sends a stronger signal than the ratification of a UN human rights treaty (Hafner-Burton et al. 2015); and submission to international election monitoring is more credible when monitors are granted unfettered access to all polling stations and allowed to publicly disclose their findings (Hyde 2012).

In order to illustrate the attractiveness of GPAs vis-à-vis alternative credibility signaling mechanisms, consider the World Bank's Doing Business Report. When the World Bank releases this report and its annual EDBI rankings, investors can immediately and simultaneously judge for themselves whether and how their baseline assumptions about the investment climate conditions of 190 jurisdictions should be updated.[10] This annual assessment does not require high levels of expert interpretation or intermediation in the same way as, say, a long-form narrative assessment (e.g. an IMF Article IV policy surveillance report). Also, whereas other credibility signaling mechanisms provide opportunities to send a one-time signal (e.g. bilateral investment treaty ratification, accession to an international organization), the Doing Business Report (and most other GPAs) allow the authorities in assessed countries to quickly and repeatedly (annually) convey the scale, scope, and the pace of domestic policy change to investors and donors.

There is some qualitative evidence that GPAs may provide a uniquely strong credibility signaling benefit. In an in-depth study of business registration reform efforts in Senegal, Gainer et al. (2016) reports that "[t]he Senegalese government considered the [Doing Business] rankings—which compared countries around the world on the basis of time, complexity, and cost of procedures like registering a business or paying taxes—to be an important signal to investors." A 2008 independent evaluation of the Doing Business Report also found that World Bank staff believe that "the aggregate ranking may motivate governments to reform because they perceive it as a signaling device for potential investors, especially foreign investors" (IEG 2008: 43).[11] It also noteworthy that the so-called "big 3" risk rating agencies—Standard and Poor's, Moody's and Fitch Group—all use a variety of GPAs to evaluate borrower creditworthiness (EITI International Secretariat 2015).

However, while credibility signaling to external actors via GPAs can provide benefits (e.g. increased economic growth and government revenue due to incoming FDI or foreign aid), it can also impose costs. Among the most important and frequently identified costs are the domestic policy autonomy

[10] There is some evidence that foreign investors do use the *Doing Business* Report in this way (Diarra 2014; Martin et al. 2014; Corcoran and Gillanders 2015). There is also evidence of other GPAs having similar information-updating impacts among donors, creditors, and investors. For example, in 2006, *Reuters* reported that the MCC's annual partner country selection process, which is based on a GPA, is "closely watched by development groups and investment firms which see it as a confidence vote in governments in the developing world" (Wroughton 2006). See also Öhler et al. 2012; Morse 2016.

[11] There are also many journalistic accounts that suggest foreign investors pay attention to the *Doing Business* Report and consider it to be a meaningful signal. For example, in 2012, the *New York Times* published a story about an investment fund manager who had decided to double down on investments in Egypt after monitoring the country's performance on the World Bank's EDBI. He said that "[o]ne of the things that attracted us to Egypt was that the World Bank had improved its 'ease of doing business' rating. ...Egypt had become more hospitable to investment. From 2004 to 2007, it had the single greatest improvement in 'ease of doing business' of any country in the world" (Luongo 2012).

losses that governments can incur when they attempt to address the performance requirements of an external actor (Pevehouse 2002; Büthe and Milner 2008; Hollyer and Rosendorff 2012). Governing elites in developing countries generally want to minimize the domestic audience and sovereignty costs of "bowing" to external pressure (Vreeland 2003; Elkins et al. 2006). GPAs are one such external pressure and there is no theoretical reason to believe that GPAs will be viewed any differently by governing elites in developing countries. Indeed, critics of GPAs charge that they often lack the detailed, country-specific diagnostic and advisory content that policymakers need (Momani 2007; Lombardi and Woods 2008; Edwards 2011) and instead encourage governments to pursue "one size fits all" policy interventions (Chhotray and Hume 2009; Goldsmith 2011). A related criticism is that GPAs create incentives to engage in de jure (or form-based) reforms rather than de facto (or function-based) reforms that are tailored to local conditions (Arruñada 2007; Hallward-Driemeier and Pritchett 2015).[12] Andrews (2013) refers to this pattern of behavior as "isomorphic mimicry" and others have warned that incentives to satisfy external demandeurs by complying with "global best practices" can divert government attention and effort away from higher priority policy challenges and limit a government's autonomy and flexibility (Soederberg 2004; Arruñada 2007; Sjöstedt 2013; Kraay and Tawara 2013; Hoey 2015; Buntaine et al. forthcoming).[13] Therefore, inasmuch as GPAs draw governing elites away from solving public problems and towards address externally-defined challenges, meeting GPA performance requirements will likely impose domestic policy autonomy costs on assessed governments.

One way that assessment suppliers can lower the (perceived or actual) costs of foregone domestic policy discretion is to involve assessed governments directly in assessment process. This participatory practice makes it easier for assessed governments to preserve their domestic policy autonomy and it encourages assessment uptake by the authorities through two different channels: legitimacy and

[12] The Government of Afghanistan's efforts to use the *Doing Business* report as a tool for informing and monitoring its business registration reform efforts provide a useful illustration of how GPAs might encourage isomorphic mimicry or efforts by assessed governments to "game" international rankings and ratings rather than solve *de facto* public problems. By most accounts, the authorities created a new business registry and reduced the cost of formal business registration in order to improve its ranking vis-à-vis other countries on the World Bank's *Doing Business* indicators (Arruñada 2007; USAID Afghanistan 2012). However, private sector development experts identify this episode of *de jure* reform as a cautionary tale, noting that while the World Bank has designated Afghanistan as a leading reformer of business registration, it remains very difficult to start and operate a new business in Afghanistan because of the need for expensive expeditors and a time-consuming licensing process that begins *after legal incorporation* (Channel 2008). IEG (2008: 28) characterizes Afghanistan's efforts as a "paper" reform, emphasizing that the Afghan authorities "simply pushed all important procedures to a stage after the legal registration of a business."

[13] Many scholars have argued that external actors (including GPA providers) should protect and preserve the policy maneuverability of decision-makers in developing countries so that they can engage in problem-driven iterative adaptation (PDIA) approaches (Pritchett al. 2012) and search for difficult-to-identify solutions that fit the local context and address *de facto* problems (Natsios 2010; Sjöstedt 2013; and Perakis and Savedoff 2015).

knowledge.[14] First, allowing assessed governments to participate in the assessment process increases the legitimacy of externally-led performance assessments and assessment processes (Andrews 2011; Swedlund 2013; Busia 2014). External evaluators often lack contextual knowledge about the government that they are responsible for assessing and fail to adequately consult with local authorities who possess significant informational advantages (Ubaldi 2013). Sourcing information from public officials in the assessed government can therefore reduce perceived bias in the assessments produced by external actors (Momani 2007; Lombardi and Woods 2008; Edwards 2011). The direct participation of the assessed government in assessment processes can also increase the probability that any final determinations that are made achieve a minimum level of resonance with public sector decision-makers in the assessed country.[15] Indeed, the evidence uptake and organizational learning literatures suggests that the most influential diagnostic and advisory tools are those that involve "the assessed" in evaluation design, data collection, and evidence interpretation processes (Patton 2008; Russ-Eft and Preskill 2009). Conversely, in cases when assessment producers fail to engage domestic policymakers and, thus, externally-generated assessments are as regarded as illegitimate and untrustworthy, local buy-in and assessment uptake by the decision-makers whom assessment providers ultimately wish to influence is less likely (Johnson et al. 2009: 389; also see Mechant-Vega and Malesky 2011).

Second, when an external evaluator involves an assessed government in its assessment process, the latter can shape the former's knowledge about its "qualities, policies, and practices" (Kelley and Simmons 2016: 20). Public officials in assessed countries are generally motivated to participate in assessment processes and eager to bring to bear local sources of knowledge because a government's reputation is a valuable asset that needs to be cultivated and protected.[16] The ability to shape knowledge and narratives about the performance of one's own government is also important because

[14] Of course, while participatory performance assessments have knowledge and legitimacy advantages, it is also possible that involving assessed entities in assessment processes could lead to capture, manipulation, and status quo bias (Rodrik 1996; Momani 2007; Swedlund 2013). However, our study is focused on explaining the determinants of assessments influence. On the broader issue of whether external assessments of government performance lead to negative, unintended consequences, see Parks et al. (2015).

[15] By way of illustration, the African Peer Review Mechanism (APRM) is officially produced by the African Union (an inter-governmental organization); however, Busia (2014: 187) characterizes the AU-sponsored African Peer Review Mechanism (APRM) process as "self-evaluation by the participating countries, validation by a group of eminent Africans, and a review of conclusions and findings by African peers." Busia (2014: 192) also argues that the "APRM methodology can be viewed as a deliberative alternative to external assessments—conducted by donors for internal corporate use—with little or no relevance for recipient partner countries in terms of deepening democratization or accountability between the state and its citizens."

[16] An extensive literature in international relations demonstrates that bad reputations and good reputations are economically and politically consequential (Downs and Jones 2002; Sartori 2005; Tomz 2007; Crescenzi 2007; Tingley and Walter 2011). For example, if a state has a reputation for default on it loans, it will be significantly more difficulty accessing international capital markets and securing new loans from external sources on favorable terms (Tomz 2007).

it allows policymakers to retain some measure of control over the nature, timing, speed, scope, and magnitude of the policy changes that assessments implicitly or explicitly advocate. Involvement in the assessment process also provides the assessed party with more evaluative information at an earlier stage in the evaluation process, thereby increasing certainty about the specific policy interventions that will most effectively address the performance requirements of a given assessment. As such, when an assessment process becomes a joint venture between the assessed and the evaluator, it should reduce the perceived risk of addressing the requirements of a performance assessment led by an external actor (Andrews 2011; Swedlund 2013; Busia 2014).

In light of these considerations, we expect that policymakers in assessed countries will weigh the credibility signaling benefits of GPAs against their policy autonomy costs and attach higher risk premia to GPAs (more specifically, cross-country benchmarking exercises) that do not allow for assessed governments to engage in the assessment process. Developing country policymakers with scarce resources and time contemplating the possibility or recalibrating their existing reform priorities and efforts in response to a cross-country benchmarking exercise need to be convinced that their actions will improve their chances of achieving the expected benefits of credibility signaling. That is to say, the authorities want stronger rather than weaker assurances that an allocation of scarce financial, human, and political resources to meet the performance requirements of a cross-country assessment will yield the expected benefits within a politically-relevant time horizon.[17] Therefore, we predict that governments will favor GPAs (more specifically, cross-country benchmarking assessments) that allow them to estimate with higher rather than lower levels of certainty the expected benefits of credibility signaling. Allowing assessed governments to participate in the assessment process performs this function. It does so by giving assessed governments more information—and reducing their uncertainty—about which policy inputs are most likely to produce the GPA improvements and reputational rewards that they wish to achieve.

To summarize, then, the key elements of our theory: we argue that the primary benefit of improving

[17] Sierra Leone's efforts to meet the performance requirements of the MCC eligibility criteria provide a useful illustration of how domestic political incentives can shape the way that governing elites respond to performance assessments (Parks 2014). In a cable dispatch from the U.S. Embassy in Freetown to State Department headquarters in Washington, the U.S. Ambassador U.S. Ambassador to Sierra Leone described a December 2008 meeting Presidential Affairs Minister Alhaji Alpha Kanu in which the minister "made several references to ... MCC [eligibility] being 'for the good of the party,' before quickly backstepping and saying 'for the good of the country'" (Perry 2008). In the cable he U.S. Ambassador offered some reflections on this meeting with the Minister: "No doubt he recognizes the political goldmine the MCC represents, and how the President and [the ruling party] would benefit if Sierra Leone enters into [an MCC] compact agreement before the 2012 elections" (Ibid.).

one's performance on a GPA is to signal one's policy credentials to foreign investors and donors; we posit also that, to achieve this benefit, the assessed government has to be willing to incur costs—specifically, the costs of limiting its own policy autonomy. Without detailed information about the size of the benefits of credibility signaling and the costs of foregone policy autonomy, we cannot know with any certainty whether benefits exceed costs (and the GPA will exert influence) or costs exceed benefits (and the GPA will fail to exert influence). Therefore, the direction of the predicted relationship between cross-country benchmarking (the defining feature of GPAs) and policy influence is ambiguous.

However, in cases where the assessed government can increase its policy autonomy (i.e. higher sovereignty and lower domestic audience costs) and reduce its uncertainty about the achievability of the net benefits (of signaling credibility to investors), we would expect the risk-adjusted cost-benefit calculation to shift in the direction of taking policy action to address the performance requirements of a GPA (cross-country benchmarking exercise). This also means that when a cross-country assessment allows for assessed governments to engage in the assessment process, it is effectively reducing the risk adjustment factor that assessed governments used in their cost-benefit calculation (or whether or not to address the performance requirements of the assessment). This risk adjustment factor, in effect, represents the imperfect information that the assessed government possesses about the likelihood of reaping the credibility signaling benefits that cross-country assessments can provide.

Based upon this theoretical reasoning, we will seek to evaluate the following testable hypotheses:

Hypothesis 1: An external assessment of government performance will exert more influence on domestic reform priorities and efforts when it allows assessed governments to participate in the assessment process.

Hypothesis 2: An external assessment of government performance that involves cross-country benchmarking will exert more influence on domestic reform priorities and efforts when it allows assessed governments to participate in the assessment process.

3. Data and Methodology

To test these hypotheses, we use a novel source of micro-level data: a survey of 3,400 government officials from 123 low-income and middle-income countries. The 2014 Reform Efforts Survey, which was conducted by the College of William and Mary's Institute for the Theory and Practice of International Relations in partnership with the National Opinion Research Center at the University of Chicago during the summer of 2014, it is particularly well suited for the purposes of our study because it allows us to quantify the policy influence of 103 different performance assessments by collecting data on the perceptions and observations of public sector decision-makers in "assessed" countries.[18] Previous studies (e.g., Pitlik et al. 2010; Andrews 2011; Parks 2014; Kelley and Simmons 2015) have examined whether, when, and how individual performance assessments have instigated policy changes in a set of countries; however, to our knowledge no study has explicitly tested which factors account for variation in policy influence across GPAs. The 2014 Reform Efforts Survey asked each respondent to evaluate the extent to which each performance assessment—with which he or she indicated some level of familiarity—influenced a government's reform priorities. This unique feature of the survey allows us to compare the differing levels of policy influence across different performance assessments.[19]

It also bears emphasizing that the data from this survey allow us to directly measure the causal phenomena of interest: whether and to what extent governments adjusted their policy priorities in response to a given GPA.[20] We asked policymakers in 126 low-income and middle-income countries—the individuals who GPA providers wish to influence—to make attributional determinations about the extent to which specific GPAs influenced specific policy priorities in specific countries and issue areas during specific periods of time.[21] Consequently, we do not rely on sophisticated econometric

[18] These survey data are particularly valuable in that they capture the experiences and observations of public sector officials from the "target" governments that assessment producers ultimately wish to influence. While surely not all survey respondents provided answers on the basis of policy decisions that they themselves made or directly witnessed, the types of respondents who participated in the survey are far more proximate to the empirical phenomena of interest (performance assessment influence on policy decisions) than any other population. A public opinion survey or a survey of experts would have been significantly less useful, as these populations generally know little about the inner-workings of government (Hoffmann-Lange 2007; Tansey 2007).

[19] See Appendix A for more information on how the sampling frame was constructed for this survey and how the survey was conducted.

[20] By contrast, the existing literature that seeks to explain the influence of GPAs usually relies on dependent variables that fall into one of two categories: those that use *de jure* and proxy variables to measure a government's "willingness to reform" (e.g. Pitlik et al. 2010: 179) and those that use outcome variables to draw inferences about a government behavior (e.g. Ölcer 2009; Öhler et al. 2012). These approaches face significant limitations. The former approach is hamstrung by the fact that many *de jure* variables and proxy variables do not accurately capture government commitment and the fact that government commitment does not necessarily result into policy change (Krasner 2009). The latter approach also suffers from an achilles' heel. Policy outcome measures are usually influence by a whole host of factors, some of which are outside the control of the government. Thus, in order to estimate whether a GPA influences a policy outcome, researchers have to rely on complex econometric techniques to infer causation that are subject to a variety of well-documented limitations.

[21] Elsewhere, we have demonstrated the construct validity of these types of elite survey-based measures of GPA influence (Parks and Davis 2017).

techniques to make causal claims in this study. We instead present prima facie evidence on the policy influence of GPAs (and the causal processes through which GPAs exert influence) from public sector decision-makers in developing countries – the ultimate "end-users" that GPA providers wish to influence.

Another attractive feature of the 2014 Reform Efforts Survey is that it provides rich data on the individual characteristics of policymakers. The survey asked each government official a battery of questions about his or her position (rank), primary area of policy expertise, job responsibilities, educational background, length of government service, familiarity with specific performance assessments, personal communications and interaction with donor agencies and international organizations, and so forth (Parks et al. 2015). This demographic information is particularly important since every GPA has a different area of focus and targeted set of end-users, and the demographic backgrounds of policymakers may have affected the ways in which they perceived or reported the influence of GPAs. Thus, the micro-level data from this survey allow us to control for these differences in the baseline characteristics of policymakers (who indicated that they were knowledgeable about one or more performance assessments) and explicitly estimate the effects of country-level and assessment-level factors on GPA influence.

Our key dependent variable of interest is the agenda-setting influence of performance assessments, as reported by government officials in 123 low-income and middle-income countries and territories. The 2014 Reform Efforts Survey asked participants (in question 31[22]) to rate the level of influence that individual performance assessments had on a given government's decision to pursue specific, participant-identified reforms in the own area of policy expertise.[23] Survey participants assigned scores on a scale from 0 (no influence at all) to 5 (maximum influence).[24] Our sample consists of 1,788

[22] The full questionnaire is provided in Parks et al. (2015).

[23] Survey participants were first asked to identify their area of policy expertise from one of the following 23 different policy domains: macroeconomic management; finance, credit, and banking; trade; business regulatory environment; investment; labor; energy and mining; infrastructure; land; decentralization; anti-corruption and transparency; democracy; public administration; justice and security; tax; customs; public expenditure management; health; education; family and gender; social protection and welfare; environmental protection; and agriculture and rural development. They also had the option to select "Foreign Policy" or "I did not have a particular area of focus", which resulted in a subsequent set of survey questions that were not specific to any one policy domain. Thus, the following questions that respondents answered about assessment influence were by and large anchored to his or her specific policy area of specialization. See Parks et al. (2015) for more details on the design and content of the *2014 Reform Efforts Survey*.

[24] The survey also asked each respondent to evaluate the influence of each assessment on the *design* of reform efforts (question 32) on the same scale of 0 (no influence) to 5 (maximum influence). We have replicated our econometric analysis using this

government officials who participated in the online questionnaire and answered a question about the relative influence of different performance assessments.[25] The unit of analysis is a government official's evaluation of each performance assessment.[26] On average, each respondent evaluated 7.3 performance assessments, yielding an N of 13,008.

We employ a multilevel model to analyze which factors impact the level of influence that an individual GPA has exerted on government reform priorities, as evaluated by government participants in the 2014 Reform Efforts Survey. The structure of our data produces three nested levels: individuals, assessments, and countries (See Figure 1). Failure to take into account the multilevel structure of a dataset such as ours "violates the assumption of independent errors and so can lead to the underestimation of the standard errors associated with contextual [or higher level] variables" (Solt 2009: 53).[27]

---

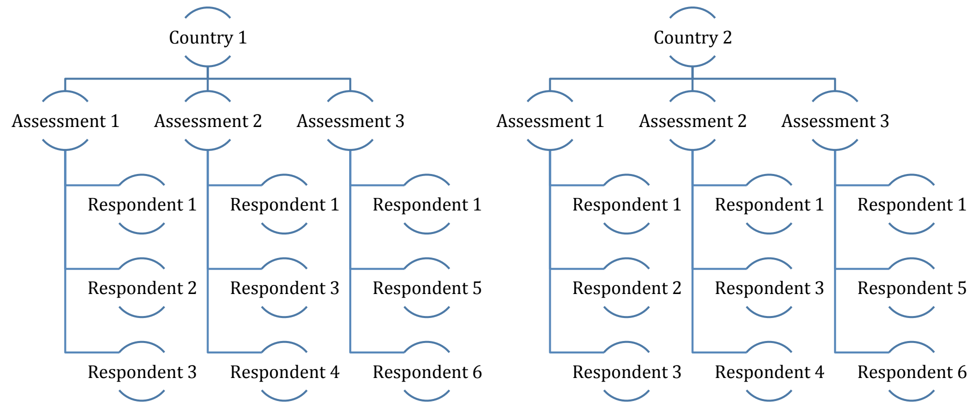
question and found that the results are almost identical to what we obtain from using question 31 (agenda-setting influence) as the dependent variable, which makes sense given a high correlation between responses in these questions (ρ>0.80).

[25] The fact that the sample for this study consists of 1,788 government officials rather than the 3,400 government officials who participated in the survey is mostly a reflection of the fact that not all government officials were knowledgeable about individual performance assessments. The *2014 Reform Efforts Survey* was designed to only allow those government officials who were knowledgeable about individual performance assessments to answer questions about their perceived level of agenda-setting influence (in question 31) and reform design influence (in question 32). Tables A-1 and A-2 in the Appendix A provide summary statistics on the characteristics of the 1,788 government officials who participated in the *2014 Reform Efforts Survey* and indicated familiarity with at least one of the individual assessments.

[26] Survey participants were first asked (in question 29) to identify a list of performance assessments with which they were familiar. They selected these assessments from a fixed list (see Parks et al. 2015 and Appendix C for more details). They were subsequently asked (in question 30) to write-in any additional performance assessments that they believed might have informed the government's reform efforts (in the respondent's particular area of policy expertise. They then individually evaluated (in question 31) the agenda-setting influence of each of these assessments on a scale of 0-5.

[27] For instance, the level of average income in a given country is a contextual characteristic specific to each country while respondent beliefs about which domestic stakeholders supported reform efforts (as well as other demographic characteristics) vary across individuals.

Figure 1: Three-Level Structure of the Survey Data



To account for potential bias in standard errors that may derive from the presence of contextual variables, we estimate a three-level model where random intercepts are introduced for each assessment-country pair and each country. Our model of assessment influence can be summarized in the following equation:

$$Inf_{ijk} = \beta_{000} + \beta_{001} AID\ DEPENDENCY_k + \beta_{002} INCOME_k + \beta_{003} POPULATION_k + \beta_{004} DEMOCRACY + \beta_{010} CROSS\_NATIONAL_{jk} + \beta_{020} GOVT\_INVOLVE_{jk} + \beta_{030} CROSS\_NATIONAL_{jk} \times GOVT\_INVOLVE_{jk} + \beta_{040} MULTILATERAL_{jk} + \beta_{050} BILATERAL_{jk} + \beta_{060} PUBLIC_{jk} + \beta_{070} PRIMARY\_DATA_{jk} + \beta_{080} PRESCRIPTIVE_{jk} + \beta_{090} INPUTS_{jk} + \beta_{100}\ SEX_{ijk} + \beta_{200} EXPERIENCE_{ijk} + \beta_{300} POLICY\ AREA_{ijk} + \beta_{400} EXECUTIVE\ SUPPORT_{ijk} + \beta_{500} LEGISLATIVE\ SUPPORT_{ijk} + \beta_{600} JUDICIAL\ SUPPORT_{ijk} + \beta_{700} CIVIL\ SUPPORT_{ijk} + \xi_{0jk} + \xi_{00k} \quad (1)$$

where i, j, and k index individuals, assessment, and country, respectively. $\boldsymbol{Inf_{ijk}}$ is a survey-based measure of assessment j's influence on reform priorities in country k, as evaluated by respondent i.

Our key independent variables relate to specific attributes of performance assessments—in particular, how they are produced and what information they contain. In order to test our hypothesis that the participation of assessed governments in assessment processes will have a positive effect on assessment influence (Hypothesis 1), we rely on a dummy variable (GOVT_INVOLVE) that takes a value of 1 when assessed governments are consulted before assessment are made public or allowed to participate in assessment production (e.g. data collection) processes; and 0 otherwise. We also use a binary variable (CROSS_NATIONAL) that measures whether or not a given performance assessment includes explicit, numerical or categorical comparisons or rankings across countries. However, we do not hypothesize that country benchmarking assessments will yield a larger or smaller policy influence dividend than country-specific assessments.[28] We instead predict (Hypothesis 2) that the CROSS_NATIONAL variable will positively affect our dependent variable only in cases where assessed governments are allowed to participate in the assessment process. In order to test this second hypothesis, we rely on an interaction effect between the CROSS_NATIONAL and GOVT_INVOLVE ($\boldsymbol{CROSS\_NATIONAL_{jk}} \times \boldsymbol{GOVT\_INVOLVE_{jk}}$). We expect this variable to register a positive effect as our theoretical prediction suggests that assessments yield a further influence dividend when government officials are consulted in the assessment process and the assessment also allows for cross-country benchmarking by explicitly ranking or rating countries.

There are of course many other assessment-level attributes that may condition assessment influence. To account for these factors, we include a battery of control variables. First, to test whether the identity of the supplier of a performance assessment affects its level of policy influence, we use two different dummy variables: one coded 1 if the assessment is produced by a multilateral institution (MULTILATERAL) and zero otherwise; and the other coded 1 if the assessment is produced by a bilateral institution (BILATERAL) and zero otherwise.[29] Second, to test whether publicly available assessments yield greater influence than those assessments to which access is restricted, we include a variable (PUBLIC) that is coded 1 if a given assessment can be accessed for free online, and zero otherwise.[30] Third, to examine whether assessments that rely on primary data are more influential than

[28] Readers will recall from section II of this paper that we have no theoretical basis to expect the credibility signaling benefits of cross-country benchmarking exercises to be larger—or smaller—than their policy autonomy costs.

[29] States may wish to delegate policy surveillance, analysis, and advisory responsibilities to multilateral institutions for several reasons (Hagen 2009; Fang and Stone 2012). They may favor supra-national delegation in order to de-politicize the assessment process by entrusting this responsibility to a third party with significant technical expertise and a higher level of perceived credibility and neutrality (Hawkins et al. 2006; Kilby 2009; Clemens and Kremer 2016). Sovereign governments face the possibility that their policy surveillance, analysis, and advisory activities will come into conflict with countervailing (geostrategic, diplomatic, or commercial) interests and pressures, thereby compromising their credibility and diminishing their policy influence vis-à-vis public sector decision-makers in the countries that are being assessed (Hawkins et al. 2006; Momani 2007). Multilateral institutions do not face this constraint, or at least not to the same extent as their bilateral counterparts. At the same time, sovereign states may wield greater assessment power than non-governmental and private organizations because they possess an outsized ability to collect reliable and useful information through their networks of embassies, partner organizations, and NGOs (Kelley and Simmons 2016).

[30] Restricting access to the analytical and advisory content of an assessment prevents the media, the general public, and various domestic political actors (e.g. trade unions, business associations, political parties) from using such information to demand policy change. Conversely, when performance assessments are made publicly available, assessment producers can equip the

those that rely exclusively on secondary data, we include a dummy variable (PRIMARY_DATA) that assumes a value of 1 if performance assessments involve the collection of primary data (e.g., through interviews, questionnaires), and zero otherwise.[31]

Fourth, in order to test whether assessments that provide more actionable (or determinate) policy guidance exert greater influence than less actionable (or determinate) assessments, we use a dummy variable (PRESCRIPTIVE) that assumes a value of 1 if an assessment provides specific guidance about how governments can improve their performance, and 0 otherwise.[32] We also use a dummy variable (INPUTS) that takes a value of 1 when an assessment measures the existence of one or more inputs (i.e., official law, policy, rule, regulation, or institution) as part of its evaluation of performance, and 0 otherwise.[33] Appendix E provides more detailed information about how these assessment-specific attributes are coded.

We also include a battery of country-level and respondent-level control variables in our models. In recognition of the fact that a government's propensity to draw upon external sources of analysis and advice may be a function of its bargaining power vis-à-vis external assessment providers (Girod and Tobin 2016), we control for net official development assistance as a percentage of gross national income (a proxy for government reliance on revenue from external assessment sponsors) (AID DEPENDENCY). To account for the fact that there may be less demand for reform in lower income countries, which may in turn condition the policy influence that external actors can exert (Andrews 2011; Loayza et al. 2012), we include a measure of per capita income (INCOME). We also include a measure of political openness (DEMOCRACY) because more liberal regimes may be more likely to

---

media, the general public, civil society organizations, and various domestic political actors with information that they need to more advocate for reform (Besley and Burges 2002; McMillan and Zoido 2004; Stoke 2001; Knecht 2010; Kelley and Simmons 2016).

[31] Aggregating and repackaging secondary data as composite indices can be attractive to assessment producers because it can lower the costs of production and reduce measurement errors (Kaufmann et al. 2007; Kraay and Tawara 2013). However, composite indices "run the risk of losing the conceptual clarity that is so crucial" if one wishes to offer specific policy recommendations to the authorities (Chowdury et al. 2008: 21; also Langbein and Knack 2010 and Thomas 2010).

[32] We include this control variable to account for the fact that performance assessments containing more actionable policy guidance may exert greater influence than performance assessments containing less actionable policy guidance. The logic of determinacy suggests that governments will favor assessments with performance requirements that are easy to understand and target with near-term policy interventions (Schimmelfennig and Sedelmeier 2004; Parks 2014; Trapnell 2011; Erkkilä 2015; Custer et al. 2016; Kelley and Simmons 2016).

[33] As Buntaine et al. (forthcoming) explain, indicators and assessments that measure (de jure) inputs tend to be more actionable than (de facto) outputs or outcomes because laws, rules, regulations, and formal institutions can be established relatively easily and in a relatively short span of time. Decisions to implement input-focused measures are also easier to pursue because they fall within the control of the executive and the legislature.

take external pressures into account because elected leaders are more reform-oriented and have a stronger need to "lock-in" reform commitments (Kapstein and Converse 2008; Montinola 2010; Kersting and Kilby 2014).[34] Additionally, we control for the size of a country's population (POPULATION) in order to account for the fact that small states also lack bargaining power (Vreeland 2003; Pop-Eleches 2009; Sharman 2009) and may need to compensate for their own limited internal administrative and technical capabilities by relying on external sources of analysis and advice (Edwards 2012).

The last set of variables accounts for baseline characteristics of survey participants. These include sex (SEX), the length of time for which a given respondent worked for the government of his/her country (EXPERIENCE), as well as his or her primary area of policy focus (POLICY AREA).[35] Also, to account for the possibility that each respondent held different beliefs about the extent to which various domestic stakeholder groups supported reform efforts, which in turn may have also impacted his or her beliefs about an assessment's level of policy influence, we control for the degree to which respondents believed reforms enjoyed support from the executive body, the legislature, the judiciary and civil society groups (EXECUTIVE SUPPORT, LEGISLATIVE SUPPORT, JUDICIAL SUPPORT, and CIVIL SUPPORT. Table D-1 in the Appendix provides descriptions of the variables included in our regression models as well as descriptive statistics.

4. Econometric Evidence on the Determinants of Assessment Influence

Table 1 presents main results from our econometric analyses. In Models 1-4, three-level models are used to estimate the effects of assessment-specific attributes on government officials' evaluations of assessment influence. Models 5 and 6 include country-fixed dummies and respondent-fixed dummies to account for variation in baseline country and respondent characteristics that cannot be easily captured by observable indicators (e.g., the propensity to cooperate with assessment producers or undertake reforms, each respondent's bias in evaluating assessments). Testing whether our

[34] Consider Turkmenistan, the country with the lowest overall ranking on our measure of agenda-setting influence. It is governed by elites who have "zero tolerance … for information that even implies a less-than-stellar performance by the government" (IREX 2014: 283).

[35] Each respondent was asked to select his or her primary policy area of focus (from 23 different policy areas) and evaluated each GPA within the policy area of focus (Parks et al. 2015).

econometric findings remain robust to the inclusion of these country and respondent fixed effects is important because, once these country or respondent characteristics are controlled for, we can confidently attribute variation in the dependent variable solely to variation in assessment-level attributes, instead of other potential country-level or respondent-level confounders.[36]

Our findings strongly support Hypothesis 1. Performance assessments are more influential when the governments that are being assessed are directly or indirectly involved in the assessment process. The effect of GOVT_INVOLVE is positive and statistically significant ($p<0.01$), which suggests that local participation in externally-led performance assessments renders the diagnostic and advisory content that is contained in such assessments more credible and useful in the eyes of policymakers in developing countries. It is also consistent with our theoretical expectation that when governing elites in developing countries will favor performance assessments that impose fewer policy autonomy costs.

Our findings also provide evidence that government officials in low- and middle-income countries find assessments that allow for numerical or categorical cross-national comparisons (CROSS_NATIONAL) are no more or less influential than those assessments that do not allow for such comparisons if the involvement of assessed governments in the assessment process is not taken into account (See Model 2). This empirical pattern is consistent with our theoretical expectation that GPAs do not on their own provide sufficiently large or certain credibility signaling benefits to convince policymakers in assessed countries to recalibrate their domestic reform priorities and effort.

Table 1: Determinants of Agenda-Setting Influence

Model	(1)	(2)	(3)	(4)	(5)	(6)
	Multi-level	Multi-level	Multi-level	Multi-level	Fixed-Effects	Fixed-Effects
GOVT_INVOLVE	0.281		0.314	0.249	0.265	0.250
	(0.030)***		(0.033)***	(0.037)***	(0.034)***	(0.030)***
CROSS_NATIONAL		-0.025	0.125	0.045	0.068	0.064
		(0.036)	(0.039)***	(0.049)	(0.046)	(0.042)
GOVT_INVOLVE×				0.180	0.130	0.127
CROSS_NATIONAL				(0.065)***	(0.063)**	(0.058)**
MULTILATERAL	0.354	0.378	0.373	0.356	0.387	0.342

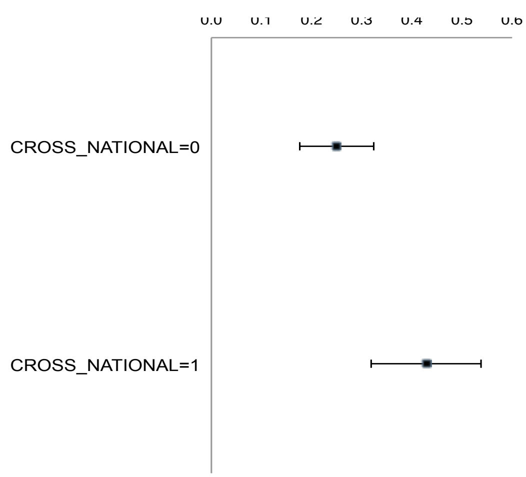
[36] For instance, it is plausible that some respondents might have found assessments that allowed for local participation in the assessment process to be more influential *not* because these assessments adopted such participatory processes, but because those respondents were in particular government positions to engage directly in the assessment process. If this was the case, it was not the attribute or design of the assessments per se that shaped the perceived influence of those assessments, but rather specific positions that respondents held in the government. In such cases, positions held by government officials are potential confounders. Including respondent-fixed effects account for such respondent-specific characteristics or experiences that could otherwise be conflated with the attributes of assessments.

	(0.044)***	(0.044)***	(0.045)***	(0.045)***	(0.046)***	(0.044)***
BILATERAL	0.157	0.065	0.192	0.166	0.208	0.168
	(0.075)**	(0.074)	(0.077)**	(0.077)**	(0.064)***	(0.060)***
PUBLIC	0.054	0.078	0.003	0.034	0.037	0.028
	(0.035)	(0.038)**	(0.039)	(0.040)	(0.038)	(0.034)
PRIMARY_DATA	0.001	0.024	-0.047	-0.069	-0.065	-0.061
	(0.032)	(0.034)	(0.035)	(0.036)*	(0.033)**	(0.031)**
PRESCRIPTIVE	-0.078	-0.045	-0.029	-0.030	-0.036	-0.032
	(0.029)***	(0.032)	(0.033)	(0.032)	(0.030)	(0.028)
INPUTS	-0.117	-0.078	-0.117	-0.123	-0.130	-0.123
	(0.033)***	(0.032)**	(0.032)***	(0.033)***	(0.029)***	(0.025)***
Respondent-Level Controls						
SEX	0.086	0.083	0.086	0.085	0.054	
	(0.080)	(0.081)	(0.080)	(0.080)	(0.074)	
EXECUTIVE	0.275	0.276	0.274	0.274	0.246	
SUPPORT	(0.069)***	(0.069)***	(0.069)***	(0.069)***	(0.068)***	
LEGISLATIVE	0.025	0.026	0.025	0.025	0.052	
SUPPORT	(0.078)	(0.078)	(0.078)	(0.078)	(0.068)	
JUDICIAL	0.295	0.294	0.294	0.293	0.298	
SUPPORT	(0.096)***	(0.096)***	(0.096)***	(0.096)***	(0.085)***	
CIVIL SUPPORT	0.265	0.261	0.265	0.266	0.281	
	(0.072)***	(0.072)***	(0.072)***	(0.072)***	(0.073)***	
Country-Level Controls						
AID	0.050	0.050	0.050	0.050		
DEPENDENCY	(0.072)	(0.073)	(0.072)	(0.072)		
INCOME	0.016	0.014	0.016	0.016		
	(0.104)	(0.105)	(0.104)	(0.104)		
POPULATION	-0.049	-0.049	-0.049	-0.049		
	(0.049)	(0.049)	(0.049)	(0.049)		
DEMOCRACY	0.010	0.009	0.011	0.011		
	(0.010)	(0.010)	(0.010)	(0.010)		
Level 3. Countries	0.442	0.441	0.442	0.442		
(Intercept Variance)	(0.037)***	(0.037)***	(0.037)***	(0.037)***		
Level 2. Assessment-	0.444	0.462	0.441	0.440		
Country Dyads	(0.036)***	(0.036)***	(0.036)***	(0.035)***		
(Intercept Variance)						
Level 1. Individual	1.269	1.269	1.269	1.269		
Responses	(0.023)***	(0.023)***	(0.023)***	(0.023)***		
(Residual Variance)						
Country Fixed Effects	No	No	No	No	Yes	No
Respondent Fixed Effects	No	No	No	No	No	Yes
N of Countries	99	99	99	99		
N of Assessment-Country Dyads	3427	3427	3427	3427		
N of Individual Responses	10504	10504	10504	10504	10504	10504

Notes: All these regressions include controls for each respondent's primary area of policy focus and length of service to the government. Models 1-4 are estimated in three-level models while Model 5 and 6 introduces country-fixed effects and respondent fixed effects. For Models 5 and 6, standard errors are clustered by respondent. All analyses use inverse-probability weights to account for variation in unit non-response rate (see Appendix B for more details). * p<0.1; ** p<0.05; *** p<0.01.

However, in the subset of cross-country assessments that allow assessed governments to participate in the assessment process, we do find a positive and statistically significant effect on agenda-setting influence. This result provides strong support for Hypothesis 2. It is consistent with our theoretical expectation that GPAs will exert policy influence when they reduce the uncertainty that public sector decision-makers in assessed countries face about the achievability of credibility signaling benefits (through participation in assessment processes with the assessment provider). Figure 2 presents the effects of GOVT_INVOLVE for the subset of assessments in our sample that involve cross-country benchmarking (CROSS_NATIONAL=1) and for subset of assessments in our sample that do not involve cross-country benchmarking (CROSS_NATIONAL=0). Whereas allowing assessed governments to participate in the assessment process increases the agenda-setting influence of the average assessment without cross-country benchmarking by 0.25, this effect size increases to 0.43 for assessments with cross-country benchmarking.

Figure 2: The Effects of GOVT_INVOLVE on Agenda-Setting Influence when CROSS_NATIONAL=0 and CROSS_NATIONAL=1



Notes: Calculated based on Model 4 in Table 1.

We also find that performance assessments from multilateral suppliers are perceived to be more influential. The effects of MULTILATERAL are positive and statistically significant ($p<0.01$) across all different model specifications (Models 1-6). On average, multilateral assessments scored 0.34-0.39 points higher (with a standard error of ±0.04-0.05) than assessments produced by non-governmental institutions and approximately 0.2 points higher than those produced by sovereign governments. Our findings also suggest that assessments produced by sovereign governments have greater policy influence than non-governmental assessments, although the effect of BILATERAL is not significant in Model 2 without controlling for the involvement of assessed governments in the assessment process.

We find little empirical support that public access to performance assessments (PUBLIC) has a positive effect on assessment influence. Nor do we find strong evidence that an assessment's reliance upon primary (as opposed to secondary) data (PRIMARY_DATA) has any consistent bearing upon its policy influence. Also, contrary to the conventional wisdom, we do not observe a strong, positive relationship between assessment actionability and assessment influence. Our first and most direct measure of assessment actionability (PRESCRIPTIVE) does not demonstrate a strong or consistent relationship with policy influence.[37] Our second and more indirect measure (INPUTS) actually registers a negative and statistically significant effect on the dependent variable, which suggests that input-focused assessments achieve less influence than those assessments that measure performance on the basis of outcomes.

Our models point to several other results. We find strong evidence for the claim that domestic political support for reform conditions assessment influence (Kahler 1992; Chassy 2014; Parks et al. 2015). Specifically, we find that reform support by the Office of the Head of State and/or Office of the Head of Government (EXECUTIVE SUPPORT) has a statistically significant and positive effect on assessment influence (p-value<0.01). These findings attest to the centrality of the executive branch in leading and shepherding (GPA-informed) reform processes. We also find that presence of reform supporters within the judicial branch (JUDICIAL SUPPORT) and domestic civil society organizations (CIVIL SUPPORT) can expand or narrow the scope of influence that performance assessments exert on reform priorities and processes. Taken together, these results point to "pass-through effects" whereby a country's propensity to undertake reform amplifies or dampens the policy influence of performance assessments (James 1996; Stone 2011; Chassy 2014). These findings also corroborate previous research that suggests external actors rely on the presence of domestic allies who endorse the policy goals—or those whom Chwieroth (2013: 267) refers to as "sympathetic domestic interlocutors"—to effectuate policy change.

[37] While statistically significant in some models, the effects of the PUBLIC, PRIMARY_DATA, and PRESCRIPTIVE variables are sensitive to model specifications and a set of control variables included on the right-hand side equation. PUBLIC and PRESCRIPTIVE are statistically significant only in Model 2 and 1, respectively.

We subjected these findings to a number of different robustness tests.[38] First, we replicate our results with an alternative outcome variable that measures the reform design influence of government performance assessments. Our results are insensitive to the inclusion of an alternative dependent variable. Second, we examine whether our results are generalizable across regions or driven by skewed regional representation within our sample. We find that our results largely hold across regions, but they are strongest in Africa. Third, recognizing the possibility that some assessment providers are also suppliers of grants and loans and respondents might ascribe influence to these providers because of their financial leverage rather than their assessment power, we control for this potential confound and we find no evidence that this attribution problem has biased our results. Fourth, to account for potential serial correlations across individual responses by assessment producer, we allow intercepts to vary by assessment producers. Our results are robust to the inclusion of these random intercepts for each producer. Fifth, we include country-fixed dummies and respondent-fixed dummies in our models to account for variation in baseline country and respondent characteristics that cannot be easily captured by observable indicators. Our results are robust to the inclusion of these country- and respondent-fixed effects.

5. Supplementary Survey Evidence on the Determinants of Assessment Influence

In summary, the econometric evidence suggests that GPAs are influential when the authorities in assessed jurisdictions have confidence that the credibility signaling benefits of policy action exceed the costs of foregone policy autonomy. However, the econometric evidence alone does not provide sufficiently comprehensive empirical support for our theoretical argument and its core assumptions. How do we know, for example, that government officials are using GPAs to signal their credibility to foreign investors or international donors? Do we have strong empirical grounds for our core assumption that government officials are balancing the credibility signaling benefits against the expected domestic policy autonomy costs of GPA-led policy action? Is there any corroborating evidence that policymakers are more confident that they will reap credibility signaling benefits when their governments are allowed to participate in the assessment process?

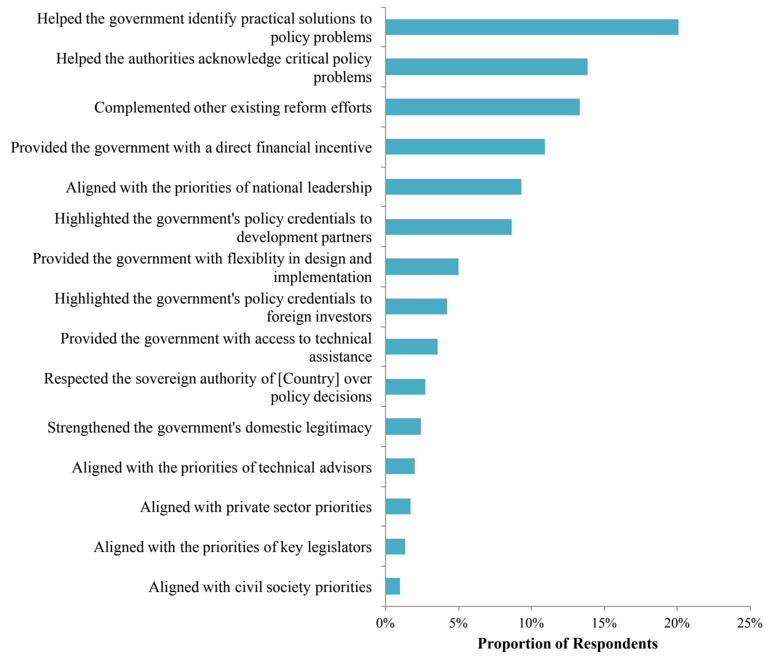
[38] These tests are described at greater length in Appendix F. Also see Models 5 and 6 in Table 1.

In order to address these questions and more fully expose our theoretical argument to empirical scrutiny, we now turn to a supplementary source of data that was collected through the 2014 Reform Efforts Survey. In cases when survey participants indicated (in question 31 and 32) that a given performance assessment achieved a minimum level of agenda-setting influence or reform design influence, they were subsequently asked to identify the primary reason why that individual assessment proved influential.[39] Respondents were given the opportunity (in question 34) to select one primary reason for assessment influence from a fixed list of fifteen options.

Across our full sample of 103 cross-country and country-specific assessments, the response option that was most frequently selected by government participants from this fixed list was that assessments provided practical solutions to help address critical issues facing the country (see Figure 4). 20% of public sector respondents identified this factor as the primary reason for assessment influence. 14% indicated that assessments exerted influence on government reform priorities and reform design efforts because they helped the government to understand or more fully appreciate the critical nature of policy problems facing the country. Another 13% of government participants indicated that assessments were influential because they promoted reform efforts that complemented existing government reform efforts. These descriptive statistics suggest that public sector decision-makers in the developing world are primarily concerned with the degree to which external assessments of government performance align with their own interests and priorities (i.e. domestic policy autonomy considerations). By contrast, credibility signaling to investors appears to be, at best, a secondary consideration across the full sample of cross-country and country-specific assessments.

[39] Participants in this survey were only asked to answer this question for performance assessments that received the highest sum of ratings in question 31 (agenda-setting influence) and 32 (design influence). When they tied in the sum, respondents were asked to answer this question only for the assessment that received a higher rating in question 31. If they still tied in question 31, one of the assessments was selected at random.

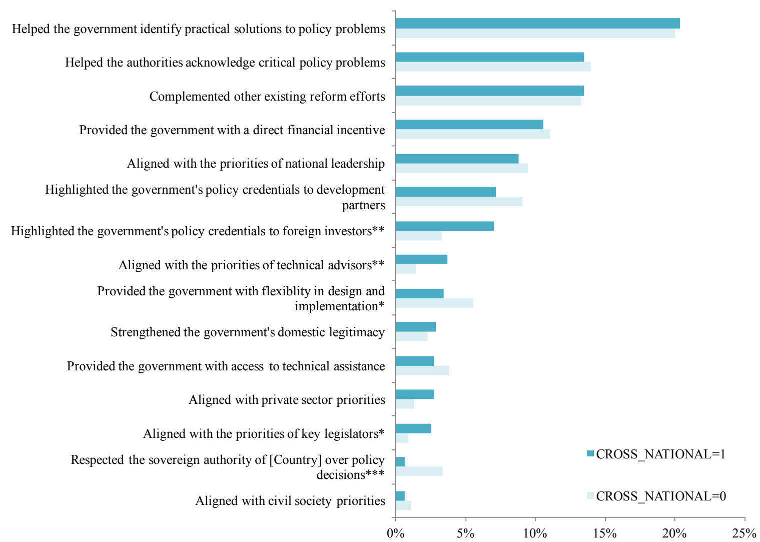
Figure 3: Primary Reasons for Assessment Influence



Notes: The figure shows the proportion of respondents who selected each response option as the primary reason for assessment influence.

Figure 4: Primary Reasons for Assessment Influence:

CROSS_NATIONAL=1 and CROSS_NATIONAL=0



Notes: The figure compares the proportion of respondents who selected each response option as the primary reason for assessment influence. The comparison is made between those assessments that allow for cross-national benchmarking (CROSS_NATIONAL=1) and assessments that do not (CROSS_NATIONAL=0). We conduct a difference-in-means test to evaluate whether the estimated difference is statistically significant. * p<0.1; ** p<0.05; *** p<0.01.

However, when we split the full sample into performance assessments that involved cross-country benchmarking and those that did not and compare differences in the proportions of respondents across these two cohorts who cited each response option, a different pattern emerges (see Figure 4). Respondent evaluations of the "reasons for assessment influence" in the cross-country benchmarking cohort were statistically more likely than the comparison group (respondent evaluations of performance assessments that did not involve cross-country benchmarking) to say that these assessments helped "the government to highlight its policy credentials to foreign investors." They were also statistically less likely to indicate that these assessments were "seen as respecting the [country's] sovereign authority over final policy decisions." These results suggest that credibility signaling to foreign investors and domestic policy autonomy are primary considerations – that cut in two different directions – for governing elites in developing countries who are considering GPA-focused policy interventions. We consider this to be strong, corroborating evidence that, when governing elites in developing countries seek to determine if it is in their interests to address the performance requirements of GPAs, they balance the benefits of credibility signaling to investors against the costs of foregone policy autonomy.

By contrast, we find no evidence that cross-country benchmarking is considered by public officials in developing countries to be a more useful device than country-specific assessments for signaling credibility to foreign donors (see Figure 4). Respondent evaluations of the "reasons for assessment influence" in the cross-country benchmarking cohort were less likely than the comparison group (respondent evaluations of performance assessments that did not involve cross-country benchmarking) to say that these assessments "created a way for the government to highlight its policy credentials to key development partners."[40] We interpret this empirical pattern as indicating that credibility signaling to development partners is not a primary reason why public officials in assessed countries consider GPAs to be influential, as compared to performance assessments that do not involve cross-country benchmarking.[41]

[40] Earlier in the survey, respondents were informed that the term "development partners" referred to "international organizations, foreign embassies, and development finance agencies."

[41] Donors have a wide array of *country-specific* assessments that they use to evaluate the performance of their partner countries (Parks et al. 2015). In cases when donors allocate aid on the basis of recipient performance (as opposed to, say, donor interests or recipient needs), they often link their financial disbursements to performance on these assessments rather than cross-country benchmarking assessments (Lu et al. 2006; Hout 2007; Savedoff 2011; Sandefur and Glassman 2015; Perakis and Savedoff 2015). Therefore, inasmuch as governing elites have incentives to signal their credibility to donors, they may prefer to use country-specific assessments instead of cross-country assessments since doing so is the arguably the most direct and certain way to maximize rewards and minimize penalties from bilateral and multilateral aid agencies and development banks. By contrast,

One reason why this may be the case is that foreign investors are more responsive than foreign donors to the policy and institutional performance of developing countries. Indeed, FDI tends to flow to countries with good credit, countries that are controlling corruption, countries that pose low risks of expropriation, and countries that are implementing economic reforms (Svensson 1998; Henisz 2000; Albuquerque 2003; Jensen 2002, 2003; Fariaa and Mauro 2009). However, aid is less responsive to these sorts of host country performance considerations (Alesina and Dollar 2000; Dollar and Levin 2006; Easterly and Williamson 2011).[42] Another potential explanation for this empirical pattern is that a large number of these country-specific assessments are designed and implemented by foreign donors for the express purpose of evaluating the credibility of current and prospective counterparts in the developing world (Parks et al. 2015), and public officials in developing countries may prefer to signal their credibility to foreign donors through the formal assessment mechanisms that donors themselves have already established. Foreign investors, by comparison, do not expect or encourage governments in developing countries to signal their credibility through country-specific assessments that they administer. Indeed, not a single country-specific assessment in our sample (see Appendix C) is produced by a foreign investor. Cross-country benchmarking exercises may therefore substitute for this absence of formal, country-specific assessment processes and create an alternative channel through which developing countries can signal their credibility to foreign investors. For both of these reasons, it is not too surprising that governing elites in developing countries have weaker incentives to use GPAs to signal to foreign donors than they do to foreign investors.]

When we split the sample into performance assessments that allowed for assessed governments to participate in the assessment process versus those that did not, we find that respondent evaluations in the former group are less likely (than respondent evaluations in the latter group) to indicate that assessments were influential because they helped the authorities better understand or appreciate the critical nature of their policy problems, and more likely to indicate that they were influential because

---

foreign investors do not produce country-specific assessments and they do not signal to actual or potential host countries that their decision to (re)invest will hinge upon whether the performance requirements of country-specific assessments are met. Instead, foreign investors tend to rely on credit rating agencies, political risk assessment firms, and suppliers of cross-country benchmarking data to help them identify places where they are likely to achieve attractive financial returns (Albuquerque 2003; Wroughton 2006; Arndt and Oman 2006; UNDP and Global Integrity 2008; Diarra 2014; Martin et al. 2014; Corcoran and Gillanders 2015; EITI 2015).

[42] The primary factors that shape the cross-country allocation of aid are donor interests (e.g. rewarding foreign policy allies and punishing foreign policy opponents) and recipient needs (e.g. numbers of people living in poverty, the severity of damage from disasters).

they promoted reform efforts that complemented existing government reform efforts (see Figure 5).[43] These findings are again consistent with our theoretical expectation that performance assessments which allow assessed governments to participate in the assessment process are less challenging of and more accommodating towards the interests and priorities of the local authorities (i.e. they impose fewer domestic policy autonomy costs).

Finally, when we split the sample of performance assessments that allowed for assessed governments to participate in the assessment process into those that also relied upon cross-national benchmarking and those that did not (see Figure 6), we find that respondent evaluations in the former group are more likely than respondent evaluations in the latter group to indicate that such assessments "created a way for the government to highlight its policy credentials to foreign investors." We also see in Figure 6 that, when we only examine the cross-country assessments that allowed for assessed governments to participate in the assessment process, survey participants do not identify respect for the assessed government's sovereignty to make final policy decisions as a reason for assessment influence. This finding lends support to our interpretation (in Section IV) of the positive and statistically significant relationship that our main interaction effect of interest ($CROSS\_NATIONAL_{jk} \times GOVT\_INVOLVE_{jk}$ from Table 1) registers on our dependent variable. Taken together, these results suggest that governing elites in developing countries are more confident that they will reap the expected benefits of credibility signaling to investors when their governments are allowed to participate in assessment processes.[44] Whereas respect for the sovereignty of the assessment government is identified as a major impediment to policy influence for cross-country assessments (i.e. cross-country assessments are, on the whole, seen as imposing significant policy autonomy costs), this impediment to influence is not identified for the smaller cohort of cross-country assessments that allow assessed governments to participate in the assessment process.[45]

---

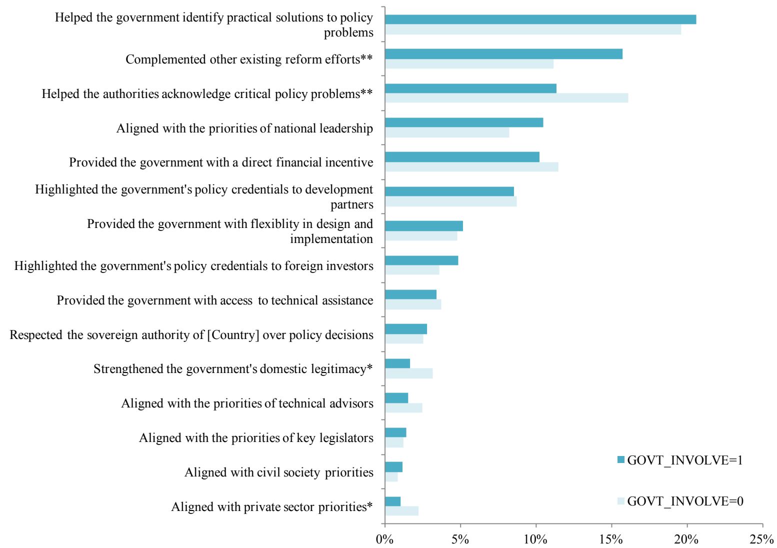
[43] See also Table G-1 in the Appendix.

[44] It is also worth noting that we find some evidence that the perceived benefits of credibility signaling benefits may increase in quantitative terms when cross-country assessments include assessed governments in the assessment process (i.e. when policy autonomy costs are lower): 7% of respondent evaluations cite credibility signaling to investors as the primary reason for assessment influence in the baseline condition of all cross-country assessments and 10% cite this reason for assessment influence in cases of cross-country benchmarking that allow assessed governments to participate. By contrast, we find that 0% of respondent evaluations cite respect for sovereignty as the primary reason for assessment influence in the baseline condition of all cross-country assessments and 0% of respondent evaluations identify this consideration as the primary reason for assessment influence in cases of cross-country benchmarking with government involvement in the assessment process.

[45] In Appendix G, we also explore whether stated reasons for assessment influence vary across different government positions types. To test if our general findings hold across different position types, we categorize government positions into three different groups–leadership positions (e.g., the head of the state or government; ministers), administrative management positions (e.g., chief of staff; directors), and programmatic and technical positions (e.g., technical specialists; program/project managers)–and

Figure 5: Primary Reasons for Assessment Influence:

GOVT_INVOLVE=1 and GOVT_INVOLVE=0



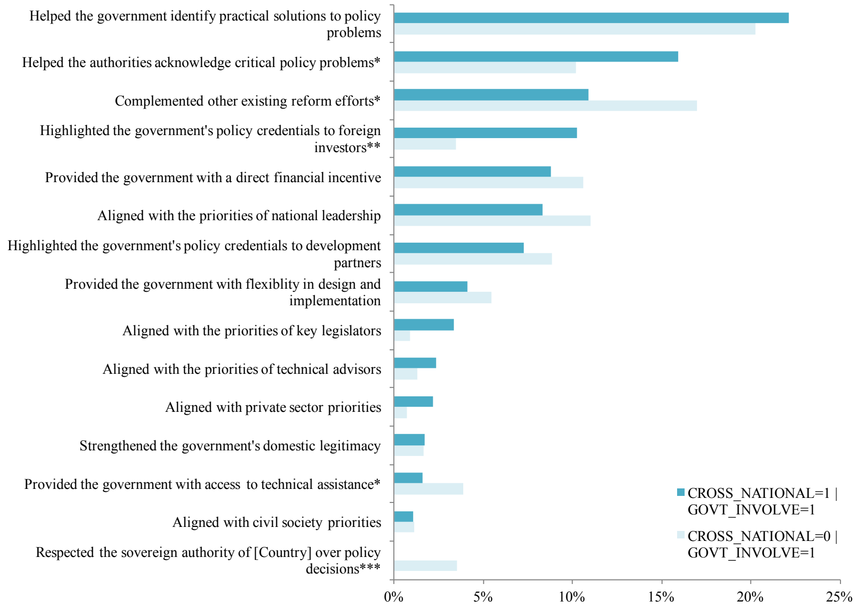
Notes: The figure compares the proportion of respondents who selected each response option as the primary reason for assessment influence. The comparison is made between those assessments that involve assessed governments in the assessment process (GOVT_INVOLVE=1) and assessments that do not (GOVT_INVOLVE=0). We conduct a difference-in-means test to evaluate whether the estimated difference is statistically significant. * p<0.1; ** p<0.05; *** p<0.01.

---

replicate this analysis of the primary reasons for assessment influence for each government position type. We find that our main results are broadly applicable across different government position types. We also find some evidence that government officials in leadership positions are more likely than government officials in technical and administrative positions to identify signaling to foreign investors and assessment alignment with national leadership priorities as primary reasons for assessment influence, which comports with our theoretical expectation that leaders in "assessed countries" must balance their interests in securing external resources and recognition while their interests in preserving domestic policy autonomy.

Figure 6: Primary Reasons for Assessment Influence:

CROSS_NATIONAL=1 and CROSS_NATIONAL=0 when GOVT_INVOLVE=1



Notes: The figure compares the proportion of respondents who selected each response option as the primary reason for assessment influence. The comparison is made between cross-country assessments that involve assessed governments in the assessment process (CROSS_NATIONAL=1 | GOVT_INVOLVE=1) and country-specific assessments that involve assessed governments in the assessment process (CROSS_NATIONAL=0 | GOVT_INVOLVE=1). We conduct a difference-in-means test to evaluate whether the estimated difference is statistically significant. * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

On balance, then, the supplementary, descriptive evidence that we gathered and analyzed from survey respondents regarding the reasons why they believe performance assessments influenced their government's reform priorities and efforts are consistent with the positive and statistically significant interaction effect that we find in our multivariate regression results.[46] It suggests that the investor signaling benefits of cross-country benchmarking are always present in the minds of policymakers, regardless of whether their government can directly engage in the assessment process (and thus receive some assurance that the government's policy sovereignty/autonomy will be respect at some minimum level). However, it also suggests that the benefits of cross-country benchmarking are not perceived to be sufficiently large or certain, relative to the costs of foregone domestic policy autonomy, unless the assessed government is allowed to engage in the cross-country assessment process.

[46] One can think of in two different ways: a shift from the "baseline condition" of assessed government participation in performance assessment processes to assessed government participation in cross-country performance assessment processes, or a shift from the baseline condition of a cross-country performance assessment to as cross-country performance assessment that allows for assessed governments to participate in the assessment process.

6. The Case of Georgia and the Doing Business Report

Thus far, we have argued that by providing numerical or categorical ratings and rankings that pit one jurisdiction against others and allow countries to signal their policy credentials to foreign investors, global performance assessments (GPAs) give the authorities in monitored and measured jurisdictions incentives to recalibrate their reform priorities and efforts. We have also argued that this effect is only "activated" when policymakers have a sufficiently high level of confidence that the benefits of credibility signaling (through GPAs) will exceed the costs of foregone policy autonomy. We have posited a particular condition under which this effect can be activated: when GPAs allow assessed governments to participate in the assessment process. Section IV and V of this study presented large-n econometric and descriptive evidence in support of this hypothesis. However, we have not yet demonstrated the plausibility of a causal mechanism that might produce this conditional effect.

For this purpose, we return to the same empirical case with which we began this study: Georgia's efforts to improve its performance on the Doing Business Report. There is little doubt that the Saakashvili administration used this GPA—a cross-country benchmarking exercise that allows assessed governments to participate in the assessment process—to guide its reform priorities and monitor its reform efforts (European Stability Initiative 2010; Bendukidze and Saakhashvili 2014). It is also clear that one of the Saakashvili administration's major motivations for pursuing such reforms was to improve the country's international reputation and attract foreign investment (European Stability Initiative 2010; Schueth 2011; De Waal 2011).[47] However, it is less well understood how the Georgian government's participation in the Doing Business assessment process increased its confidence that the implementation of specific policy actions would produce the credibility signaling benefits that it hoped to achieve.

Unlike many other GPAs, the Doing Business report allows the governments that it evaluates to participate in its annual assessment process. The World Bank team responsible for the Doing Business report has put in place various safeguards to protect again statistical fraud and make it difficult for assessed countries to "game" their rankings (Gregory 2012; Besley 2015); however, its does allow assessed governments to contribute primary data that can inform and influence the World Bank's

[47] Kakha Bendukidze, the country's State Minister of Reform Coordination, seized upon the cross-country rankings in the World Bank's annually-generated *Doing Business Report* as a uniquely useful means by which the government could signal its credibility to foreign investors (European Stability Initiative 2010; Schueth 2011; De Waal 2011).

annual ranking exercise (IEG 2008). The World Bank also engages with governments prior to the release of their annual GPA to advise and assist those that who wish to make policy changes that will result in improved global rankings (IEG 2008).

The Saakashvili administration was aware of these unique features of the Doing Business report and aggressive in its efforts to engage with the World Bank's Doing Business Team. It independently generated domestic regulatory burden data that the World Bank could use to inform its ranking of the country in future assessment rounds (USAID 2007; Schueth 2012).[48] Officials from Georgia's Office of the State Minister on Reforms Coordination, National Agency of Public Registry, and the Ministry of Economic Development were also asked by the World Bank to serve as annual "contributors" to its Doing Business Report, effectively making Georgian government officials local data collectors for the World Bank.[49] A 2007 report produced by the USAID contractor that worked closely with the Office of the State Minister of Reform Coordination on EDBI-focused reforms suggests that these local data collection efforts were consequential: "Georgia's standing in the 2007 [Doing Business report] will be even better than it might have been because the World Bank was so impressed with [our] methodology and baseline results … that they accepted the more accurate data we collected."[50] In this regard, the direct involvement of the Government of Georgia in the assessment process created opportunities for the assessed government to influence the GPA provider and the country's ranking on the GPA.

But this collaborative relationship between the assessment provider and the assessed government led to influence cutting in both directions. Georgia's efforts to improve its performance on the subset of Doing Business indicators that measure the ease of closing a business are a case in point. Simeon Djankov (the creator of the Doing Business Report and then leader of the World Bank's Doing Business Team) and Kakha Bendukidze (Georgia's State Minister of Reform Coordination) co-drafted bankruptcy legislation in 2007 that would improve Georgia's performance on these sub-component indicators of

[48] They did so with support from USAID's Business Climate Reform project (USAID 2009).

[49] According to a 2008 report from the "*Doing Business* [DB] creates its annual country rankings using information supplied by persons deemed knowledgeable about selected laws and regulations in each country covered. The DB team identifies individual lawyers, notaries, officials, and firms and requests that they provide information on one or more specified DB topics" (IEG 2008: 13). The same report notes that approximately 18% of the contributors to the *Doing Business* Report are government officials from assessed countries.

[50] We are not suggesting that the Georgian government engaged in statistical fraud or otherwise compromised the credibility of the World Bank's assessment. Rather we are arguing that the *Doing Business* Report and other GPAs that allow for or encourage involvement of the assessed governments in the assessment process are characterized by evaluative information traveling in two directions: from the assessment provider to the assessed government, and from the assessed government to the assessment provider.

the EDBI (Schueth 2007). This attempt by the assessment producer and the assessed government to jointly design reform efforts almost certainly increased the confidence of the Georgian policymakers that their inputs (the design and passage of a new law) would yield the expected output (an increase on the EDBI). Indeed the assessment provider provided the assessed government with such assurances. It noted in its Doing Business 2008 report that Georgia "passed a new [bankruptcy] law [which] maximizes the value of debtors' assets, sets shorter time limits, regulates bankruptcy trustees and strengthens creditors' rights. In place of a liquidation process that takes 3.5 years on average, the law establishes bankruptcy procedures that should take less than 1 year in the event of reorganization and just 6 months if the business is slated for liquidation. That would allow Georgia to enter the top 10 list on the speed of resolving bankruptcy" (World Bank and IFC 2007, 55, emphasis added).

This empirical case suggests that it is plausible that the conditional effect identified in Sections IV and V operates through an increase in information and a reduction in uncertainty. That is to say, if governing elites in developing countries are weighing credibility signaling benefits and domestic policy autonomy costs under conditions of uncertainty, they will likely favor those GPAs that provide greater confidence that a given policy intervention (or set of policy interventions) will yield the expected net benefits. More specifically, when a cross-country assessment allows for assessed governments to engage in the assessment processes, it likely reduce the risk adjustment factor that the assessed governments use in their cost-benefit calculation of whether or not to address the performance requirements of the assessment. Our central argument is that when this risk adjustment factor rises, the prospects for GPA influence decline; conversely, when this risk adjustment factor falls, the prospects for GPA influence increase.

7. Conclusions

The dawn of the new millennium has coincided with the dawn of a new empirical phenomenon in world politics: the production, promotion, and use of global performance assessments (GPAs). Sovereign governments, multilateral institutions, and non-governmental organizations have invented a vast array of assessment tools—including conditional aid and debt relief programs, blacklists and watch lists, organizational accession procedures, cross-country benchmarking initiatives, and in-depth, country-specific diagnostics—to evaluate and measure the performance of countries across many different policy domains. This new empirical phenomenon is both under-researched and poorly understood.

GPAs have different aims, design features, methods of measurement, and substantive areas of focus, but they all in some way seek to influence the policy decisions of governments (Pitlik et al. 2010; Parks et al. 2015; Kelley and Simmons 2016). Some increase the financial or reputational benefits that policymakers can reasonably expect to achieve if they alter their policy behavior. Others socially sanction or impose financial penalties on those who postpone or rollback reforms. Still others seek to empower change agents with new sources of data, analysis, and advice.

Scholars and policymakers generally agree that GPAs can hasten the adoption and implementation of policy, legal, regulatory, and institutional changes (Noland 1997; Schimmelfennig and Sedelmeier 2004; Parks 2014; Kelley and Simmons 2015). However, there is little systematic evidence about the conditions under which these performance assessments instigate actual changes in policy behavior.

There are a number of studies that seek to answer these questions for individual performance assessments (Noland 1997; Pitlik et al. 2010; Parks 2014; Kelley and Simmons 2015). However, to our knowledge, there is no systematic, large-n research that accounts for why some performance assessments have more influence on the reform priorities and efforts of some governments than others. We have attempted to correct this problem by drawing upon a survey of approximately 1,800 government officials that provided comparative data on the agenda-setting and reform design influence of 103 performance assessments in 123 low- and middle-income countries and territories.

With this novel source of elite survey data, we tested a number of hypotheses about the conditions under which performance assessments influence the reform priorities and reform design decisions of public sector decision-makers in developing countries. Our econometric analysis suggests that cross-national benchmarking exercises (one of the key definitional criterion for whether a given external assessment of government performance is a GPA) are more influential than country-specific assessments because they allow states to efficiently and precisely differentiate their performance from the performance of other states, and thereby transmit stronger credibility signals to foreign investors (than they would otherwise be able to transmit through the traditional, external mechanisms of policy surveillance and assessment). However, we find that this "GPA advantage" only obtains under a specific condition: when assessed governments are given the opportunity to engage in the assessment process. We argue that this is the case because government participation in cross-country assessment processes reduce the high levels of uncertainty that policymakers generally face regarding the achievability of the credibility signaling benefits that GPAs can provide.

When these econometric findings are considered alongside corroborating qualitative evidence from government officials in 123 assessed countries (who were asked in the 2014 Reform Efforts Survey to identify the primary reasons why individuals GPAs influenced their government's reform priorities and reform design decisions between 2004 and 2013), we argue that a coherent causal narrative emerges: by providing numerical or categorical ratings and rankings that pit one jurisdiction against others and allow countries to signal their policy credentials to foreign investors, cross-country benchmarking exercises give the authorities in monitored and measured jurisdictions incentives to improve their performance. However, this effect is only "activated" when policymakers have a sufficiently high level of confidence that the benefits of credibility signaling (through GPAs) will exceed the costs of foregone policy autonomy. Thus, the likelihood that a GPA's diagnostic and advisory content will influence the reform priorities and efforts of those whom it evaluates is conditional upon its ability to reduce policy autonomy costs and increase the certainty among policymakers that the credibility signaling benefits that they hope to achieve are in fact achievable.

The composite picture that emerges is not one of governments being cajoled or coerced by external actors into pursuing reforms that they would not otherwise wish to pursue. Rather, our results suggest that the domestic resonance and credibility of external analysis and advice is higher when it is co-generated with "the evaluated," and that assessment suppliers can increase their influence by incorporating contextual information and insights from the domestic authorities. In this regard, our findings puncture the conventional wisdom that one needs to create a firewall between evaluators and those who they evaluate to preserve the independence, integrity, and credibility of the process. Influential assessment providers are evidently those who treat the governments they evaluate as partners in a collaborative process (that requires local buy-in) rather than subjects in an arms-length evaluation.

Our results also speak to the question of whether GPAs represent a new empirical phenomenon in world politics or simply "old wine in new wineskins" (since states, IOs, and NGOs engaged in policy surveillance and assessment activities long before GPAs existed). Only 35 of the 103 assessments in our sample involve cross-country performance benchmarking; the remaining 68 assessments more closely resemble the traditional policy surveillance and assessment activities that existed prior to the arrival of GPAs. We have exploited this source of variation and demonstrated that GPAs have greater policy influence—or "assessment power"—than more traditional forms of external surveillance and assessment that do not involve cross-country benchmarking.

References

Abiad, Abdul and Ashoka Mody. 2005. Financial Reform: What Shakes It? What Shapes It? American Economic Review 95: 66–88.

Albuquerque, Rui. 2003. The composition of international capital flows: risk sharing through foreign direct investment. Journal of International Economics 61 (2): 353-383.

Alesina, Alberto and David Dollar. 2000. "Who Gives Foreign Aid to Whom and Why?" Journal of Economic Growth 5(1): 33-63.

Andersson, Staffan, and Paul Heywood. 2009. The politics of perception: Use and abuse of Transparency International's approach to measuring corruption. Political Studies 57 (4): 746–767.

Andrews, Matt. 2011. Which Organizational Attributes Are Amenable to External Reform? An Empirical Study of African Public Financial Management. International Public Management Journal 14(2): 131–56.

Andrews, Matt. 2013. The Limits of Institutional Reform in Development: Changing Rules for Realistic Solutions. Cambridge, UK: Cambridge University Press.

Andrews, Matt. 2015. Explaining Positive Deviance in Public Sector Reforms in Development. World Development 74: 197-208.

Arias, Eric, James R. Hollyer and B. Peter Rosendorff. 2014. Leader Survival, Regime Type and Bilateral Investment Treaties. Working Paper.

Arndt, C. 2008. The politics of governance ratings. International Public Management Journal 11 (3): 1-23

Arndt, Christiane and Charles Oman. 2006. Uses and Abuses of Governance Indicators. Paris: OECD.

Arruñada, Benito, 2007. Pitfalls to avoid when measuring institutions: Is Doing Business damaging business? Journal of Comparative Economics 35(4): 729-747.

Asian Development Bank (ADB). 2005. Multilateral Development Bank Technical Meeting on Performance Based Allocation Methods. Meeting at AsDB Headquarters on 24-25 January 2005. Manila: ADB.

Bartels, Frank L., Francesco Napolitano, and Nicola E. Tissi. FDI in sub-Saharan Africa: A longitudinal perspective on location-specific factors (2003-2010). International Business Review 23: 516-529.

Beaman L, Duflo E, Pande R, & Topalova, P. 2010-11. Political reservation and substantive representation: evidence from Indian village councils. India Policy Forum, Brookings and NCAER. 7: 159–191.

Bendukidze, Kakha and Mikheil Saakhashvili. 2014. Georgia The Most Radical Catch-Up Reforms, In Anders Aslund and Simeon Djankov, eds. The Great Rebirth: The Victory of Capitalism over Communism, Peterson Institute for International Economics, November.

Besley, Timothy. 2015. Law, regulation, and the business climate: the nature and influence of the World Bank Doing Business project. Journal of Economic Perspectives 29 (3): 99-120.

Besley, Timothy and Robin S. L. Burgess. The Political Economy of Government Responsiveness: Theory and Evidence from India. Quarterly Journal of Economics 117 (4): 1415-1451.

Biglaiser, Glen and Karl DeRouen. 2010. The Effects of IMF Programs on U.S. Foreign Direct Investment in the Developing World. The Review of International Organizations 5(1): 73-95.

Brooks, Sarah, Raphael Cunha, and Layna Mosley. 2015. Categories, Creditworthiness and Contagion: How Investors' Shortcuts Affect Sovereign Debt Markets. International Studies Quarterly 59 (3) 587-601.

Bueno de Mesquita, Bruce, and Alastair Smith. 2010. Leader Survival, Revolutions and the Nature of Government Finance. American Journal of Political Science 54: 936-950.

Buntaine, Mark, Bradley Parks, and Benjamin Buch. forthcoming. Aiming at the Wrong Targets: The Domestic Consequences of International Efforts to Build Institutions. International Studies Quarterly.

Busia, Kojo. 2014. "The Search for Domestic Accountability in Development Aid: What Role for the APRM in Reshaping Governance in Africa?" Problems, Promises, and Paradoxes of Aid: Africa's Experience, edited by Muna Ndulo and Nicolas van de Walle. Pp. 171-210. Newcastle upon Tyne, UK: Cambridge Scholars Publishing/Institute for African Development.

Büthe, Tim, and Helen V. Milner. 2008. "The Politics of Foreign Direct Investment into Developing Countries: Increasing FDI through International Trade Agreements." American Journal of Political Science 52 (October): 741–62.

Channell, Wade. 2008. Uses and Abuses of Doing Business Indicators. Washington DC: USAID.

Chassy, Aaron. 2014. "Civil Society and Development Effectiveness in Africa." In Problems, Promises, and Paradoxes of Aid: Africa's Experience. Newcastle upon Tyne: Cambridge Scholars Publishing.

Chhotray, Vasudha and David Hulme. 2009. Contrasting Visions for Aid and Governance in the 21st Century: The White House Millennium Challenge Account and DFID's Drivers of Change. World Development 37(1): 36-49.

Chowdhury, Afroza, Nathaniel Heller, and Jonathan Werve. 2008. A User's Guide to Measuring Corruption. Edited by Marie Laberge and Joachim Nahem. UNDP

Chwieroth, Jeffrey. 2007a. Neoliberal economists and capital account liberalization in emerging markets. International Organization 61 (2): 443-463.

Chwieroth, Jeffrey. 2007b. Testing and Measuring the Role of Ideas: The Case of Neoliberalism in the International Monetary Fund. International Studies Quarterly 51 (1): 5-30.

Chwieroth, Jeffrey M. 2013. The Silent Revolution: How the Staff Exercise Informal Governance over IMF Lending. Review of International Organizations 8(2): 265-90.

Corcoran, Adrian and Robert Gillanders. 2015. Foreign direct investment and the ease of doing business. Review of World Economics 151 (1): 103-126.

Clemens, Michael and Michael Kremer. 2016. The New Role for the World Bank. Journal of Economic Perspectives 30 (1): 53–76.

Crescenzi, Mark. 2007. "Reputation and International Conflict." American Journal of Political Science 51: 382-396.

Custer, Samantha, Takaaki Masaki, Tanya Sethi, Rebecca Latourell, Zachary Rice, and Bradley Parks. 2016. Governance Data: Who Uses It and Why? Williamsburg, VA: AidData and the Governance Data Alliance.

Dahl-Østergaard, Tom, Sue Unsworth, Mark Robinson, and Rikke Ingrid Jensen. 2005. Lessons Learned on the Use of Power and Drivers of Change Analyses in Development Cooperation. Review commissioned by the OECD DAC Network on Governance.

David-Barrett, Elizabeth and Ken Okamura. 2016. Norm Diffusion and Reputation: The Rise of the Extractive Industries Transparency Initiative. Governance 29 (2): 227–246.

De Waal, Alex. 2011. Georgia's Choices: Charting a Future in Uncertain Times. Washington DC: Carnegie Endowment for International Peace.

Dhaliwal, Iqbal and Caitlin Tulloch. 2012. From research to policy: using evidence from impact evaluations to inform development policy. Journal of Development Effectiveness 4 (4): 515-536.

Dhonte, Pierre. 1997. Conditionality as an Instrument of Borrower Credibility. IMF Paper on Policy Analysis and Assessment 97/2. Washington: International Monetary Fund.

Di Puppo, Lili. 2014. The Construction of Success in Anti-Corruption Activity in Georgia. East European Politics 30(1): 105–22.

Diarra, Gaoussou. 2014. How far do foreign financings influence doing business in developing countries? Paper presented at the Doing Business: Past, Present and Future of Business Regulation Conference at Georgetown University's McDonough School of Business. Washington, DC. February 20-21, 2014. Accessed at http://www.doingbusiness.org/special-features/conference/~/media/GIAWB/Doing%20Business/Documents/Miscellaneous/Conference2014/S3-3.pdf

Dinmore, Guy. 2005. Georgia calls on Europe for coherent policy. The Financial Times. September 25, 2005.

Dollar, David and Victoria Levin. 2006. The increasing selectivity of foreign aid, 1984–2003. World Development 34(12): 2034–2046.

Downs, George and Michael Jones. 2002. Reputation, Compliance, and International Law. Journal of Legal Studies 31: S95-S114.

Dreher, Axel and Christoph Moser. 2010. Do Markets Care About Central Bank Governor Changes? Evidence from Emerging Markets. Journal of Money, Credit, and Banking 42 (8): 1589-1612.

Dreher, Axel and Stefan Voigt. 2011. Does membership in international organizations increase governments' credibility? Testing the effects of delegating powers. Journal of Comparative Economics 39 (3): 326-348.

Dreher, Axel, Peter Nunnenkamp, and Hannes Öhler. 2012. Why it pays for aid recipients to take note of the Millennium Challenge Corporation: Other donors do! Economics Letters 115: 373-375.

Dreher, Axel, Heiner Mikosch and Stefan Voigt. 2010. Membership Has Its Privileges - the Effect of Membership in International Organizations on FDI. CESifo Working Paper Series, (3231).

Dreher, Axel, Silvia Marchesi, and James Vreeland. 2008. The political economy of IMF forecasts. Public Choice 137(1): 145–171.

Easterly, William. And Williamson, Claudia R. 2011. Rhetoric vs. reality: The best and worst of aid agency practices. World Development 39: 1930-1949.

Edwards, Martin S. 2011. Under the Microscope: Some Findings from the 2011 Triennial Surveillance Review. Written for the New Rules for Global Finance Coalition, December 1, 2011.

Edwards Jr. D. Brent. 2012. Small states and big institutions: USAID and education policy formation in El Salvador. Current Issues in Comparative Education 15 (1): 82-98.

Elkins, Zachary, Andrew T. Guzman, and Beth A. Simmons. 2006. Competing for capital: The diffusion of bilateral investment treaties, 1960-2000. International Organization 60 (4): 811-846.

Elkins, Zachary, Andrew T. Guzman, and Beth A. Simmons. 2008. "Competing for Capital: The Diffusion of Bilateral Investment Treaties, 1960-2000." In The Global Diffusion of Markets and Democracy. Cambridge: Cambridge University Press.

Erkkilä, Tero. 2015. Global Governance Indices as Policy Instruments: Actionability, Transparency and Comparative Policy Analysis. Journal of Comparative Policy Analysis: Research and Practice.

Extractive Industries Transparency Initiative (EITI) International Secretariat. 2015. Quantifying Intangibles Using the Extractive Industries Transparency Initiative (EITI) in Credit Rating Assessments. Oslo, Norway: EITI International Secretariat.

Fang, Songying and Erica Owen. 2011. International Institutions and Credible Commitment of Non-Democracies. The Review of International Organizations 6(2): 141-62.

Fang, Songying, and Randall Stone. 2012. International Organizations as Policy Advisors. International Organization 66 (4): 537-71.

Fariaa, André, and Paolo Mauro. 2009. Institutions and the external capital structure of countries. Journal of International Money and Finance 28 (3): 367-391.

Flores, Thomas, Gabriella Lloyd and Irfan Nooruddin. 2015. The Technocratic Advantage: Does Technocratic Leadership Attract International Financing? Working Paper.

Frot, E. and J. Santiso 2010. Crushed aid: fragmentation in sectoral aid. OECD Working Paper No. 284. Paris, France: Organisation for Economic Co-operation and Development.

Gainer, Maya, Stefanie Chan, and Laura Skoet. 2016. Faster Together: A One-Stop Shop for Business Registration in Senegal, 2006–2015. Princeton: Innovations for Successful Societies.

Garriga, Ana Carolina, and Brian J. Phillips. 2013. Foreign Aid as a Signal to Investors Predicting FDI in Post-conflict Countries. Journal of Conflict Resolution 58 (2): 280-306.

Girod, Desha M. and Jennifer Tobin. 2016. Take the Money and Run: The Determinants of Compliance with Aid Agreements. International Organization 70 (1): 209-239.

Goldsmith, Arthur A. 2011. No Country Left Behind? Performance Standards and Accountability in US Foreign Assistance. Development Policy Review 29: s157–s176.

Gray, Julia. 2009. International Organization as a Seal of Approval: European Union Accession and Investor Risk. American Journal of Political Science 53 (4): 931-949.

Gray, Julia. 2013. The Company States Keep. International Economic Organization and Sovereign Risk in Emerging Markets. Cambridge, UK: Cambridge University Press.

Gregory, Neil. 2012. Doing Business and Doing Development. World Economics 13 (3): 13-26.

Grindle, Merilee. 2004. Despite the Odds: The Contentious Politics of Education Reform. Princeton, NJ: Princeton University Press

Grossman, Herschel I. 1980. "Rational Expectations, Business Cycles, and Government Behavior." In Rational Expectations and Economic Policy. Edited by Stanley Fischer. Chicago: University of Chicago Press.

Hafner-Burton, Emilie M. 2008. Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem. International Organization 62 (4): 689-716.

Hafner-Burton, Emilie M., Edward D. Mansfield and Jon C.W. Pevehouse. 2015. Human Rights Institutions, Sovereignty Costs and Democratization. British Journal of Political Science 45 (1): 1-27.

Hollyer, James R. and B. Peter Rosendorff. 2012 Leadership Survival, Regime Type, Policy Uncertainty and PTA Accession. International Studies Quarterly 56(4): 748–764.

Hollyer, James R., B. Peter Rosendorff, and James Raymond Vreeland. 2011. Democracy and Transparency. Journal of Politics 73 (4): 1191-1205.

Hagen, Rune Jansen. 2009. Basic Analytics of Multilateral Lending and Surveillance. Journal of International Economics 79 (2009):126-13.

Hallward-Driemeier, Mary and Lant Pritchett. 2015. How Business is Done in the Developing World: Deals versus Rules. Journal of Economic Perspectives 29 (3): 121–140.

Hawkins, Darren G., David A. Lake, Daniel L. Nielson, and Michael J. Tierney. 2006. Delegation and Agency in International Organizations. Cambridge, UK: Cambridge University Press.

Henisz, Witold J, 2000. The Institutional Environment for Multinational Investment. Journal of Law, Economics and Organization16 (2): 334-64.

Hoffmann-Lange, Ursula. 2007. "Methods of Elite Research." In Oxford Handbook of Political Behavior, edited by Russell J. Dalton, Hans-Dieter Klingemann. Pp. 910-927. Oxford, UK: Oxford University Press.

Hoey, Lesli. 2015. "Show me the Numbers": Examining the Dynamics Between Evaluation and Government Performance in Developing Countries. World Development 70: 1–12.

Hout, Wil. 2007. The Politics of Aid Selectivity Good Governance Criteria in World Bank, U.S. and Dutch Development Assistance. London, UK: Routledge.

Hyde, Susan. 2012. "Why Believe International Election Monitors" in Credibility and Non-Governmental Organizations in a Globalizing World: When Virtue is Not Enough, edited by Peter Gourevitch, David A. Lake, and Janice Gross Stein. Cambridge, UK: Cambridge University Press.

Hyde, Susan. 2015. The Problem of Insincere Compliance in International Relations: Norms, Policy Diffusion and International Expectations. Working Paper.

Hyden, Goran. 2008. After the Paris declaration: Taking on the issue of power. Development Policy Review 26 (3): 259–74.

Independent Evaluation Group of The World Bank (IEG). 2008. Doing Business: An Independent Evaluation. Taking the Measure of The World Bank-IFC Doing Business Indicators. Washington D.C.: World Bank. Retrieved from http://siteresources.worldbank.org/EXTDOIBUS/Resources/db_evaluation

International Monetary Fund (IMF). 2006. Evaluation Report: Financial Sector Assessment Program. Washington DC: Independent Evaluation Office (IEO) of the International Monetary Fund. Retrieved from http://www.ieo-imf.org/ieo/files/completedevaluations/FSAP_Main_Report.pdf

International Monetary Fund (IMF). 2007. Evaluation Report: IMF Exchange Rate Policy. Washington DC: Independent Evaluation Office (IEO) of the International Monetary Fund. Retrieved from http://www.ieoimf.org/ieo/files/completedevaluations/05172007exrate_full.pdf

IREX. 2014. Media Sustainability Index 2014: The Development of Sustainable Independent Media in Europe and Eurasia. Washington DC: IREX. Retrieved from http://www.irex.org/sites/default/files/u105/EE_MSI_2014_Turkmenistan.pdf

Jacoby, Wade. 2006. Inspiration, Coalition, and Substitution: External Influences on Postcommunist Transformations. World Politics 58 (4): 623-651.

James, Harold. 1996. International Monetary Cooperation since Bretton Woods. Washington, DC: IMF.

Jensen, Nathan M. 2002. "Economic Reform, State Capture, and International Investment in Transition Economies." Journal of International Development 19: 821–824.

Jensen, Nathan M. 2003. "Democratic Governance and Multinational Corporations: Political Regimes and Inflows of Foreign Direct Investment." International Organization 57(3): 587-616.

Johnson, Kelli, Lija O. Greenseid, Stacie A. Toal, Jean A. King, Frances Lawrenz, and Boris Volkov. 2009. Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005. American Journal of Evaluation 30 (3): 377-410.

Kahler, M. 1992. External Influence, Conditionality, and the Politics of Adjusmtent. In S. Haggard & R. Kaufmann (Eds). The Politics of Economic Adjustment (pp. 89-133). Princeton, NJ: Princeton University Press.

Kapstein, Ethan B. and Nathan Converse. 2008. The Fate of Young Democracies. Cambridge, UK: Cambridge University Press.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2005. Governance Matters IV: Governance Indicators for 1996-2004. World Bank Policy Research Working Paper 3630. Washington D.C.: World Bank.

Kaufmann, Daniel, Aart Kraay and Massimo Mastruzzi. 2007. Growth and Governance: A Reply. The Journal of Politics 69 (2): 555-562.

Keefer, Philip. 2007. Clientelism, Credibility, and the Policy Choices of Young Democracies. American Journal of Political Science 51 (4): 804-821.

Kelley, Judith G. 2004. Ethnic Politics in Europe: The Power of Norms and Incentives. Princeton, New Jersey: Princeton University Press.

Kelley, Judith G. and Beth A. Simmons. 2015. Politics by Number: Indicators as Social Pressure in International Relations. American Journal of Political Science 59 (1): 55-70.

Kelley, Judith G. and Beth A. Simmons. 2016. Introduction: Global Assessment Power in the Twenty-First Century. Paper prepared for the Annual Meeting of the American Political Science Association. Philadelphia, PA. September 1-4, 2016.

Kerner, Andrew, Morten Jerven, and Alison Beatty. 2015. Does it Pay to be Poor? Testing for Systematically Underreported GNI Estimates. The Review of International Organizations (December 2015): 1-38.

Kersting, Erasmus and Christopher Kilby. 2014. Aid and Democracy Redux. European Economic Review 67: 125-143.

Kilby, Christopher. 2009. The Political Economy of Conditionality: an Empirical Analysis of World Bank Loan Disbursements. Journal of Development Economics 89 (1): 51-61.

Knecht, Thomas. 2010. Paying Attention to Foreign Affairs: How Public Opinion Affects Presidential Decision Making. University Park, PA: Penn State University Press.

Kraay, Aart and Norikazu Tawara. 2013. Can Specific Policy Indicators Identify Reform Priorities? Journal of Economic Growth 18 (3): 253-28.

Krasner, Stephen D. 2009. State Development and Foreign Aid: Modernization Theory, Institutional Capacity, and Rational Choice Institutionalism Meet Foreign Policy. Unpublished Manuscript. Palo Alto, CA: Stanford University.

Krasner, Stephen D. 2011. Foreign Aid: Competing Paradigms. Journal of Intervention and Statebuilding 5 (2): 123-149.

Kydd, Andrew. 2000. “Trust, Reassurance, and Cooperation.” International Organization 54 (2): 325-357.

Lake, David, and Christopher J. Farris. 2014. Why International Trusteeship Fails: The Politics of External Authority in Areas of Limited Statehood. Governance 27 (4): 569–587.

Langbein, Laura and Steve Knack. 2010. The World Governance Indicators: Six, One, or None? Journal of Development Studies 46 (2): 350-370.

Lawson, Andrew, Richard Gerster, and David Hoole. 2005. Learning from Experience with Performance Assessment Frameworks for General Budget Support. November 2005. Synthesis Report. Swiss State Secretariat for Economic Affairs. Retrieved from http://www.gersterconsulting.ch/docs/synthesis_paf_report.pdf

Lebovic, James H. and Erik Voeten. 2009. The Cost of Shame: International Organizations and Foreign Aid in the Punishing of Human Rights Violators. Journal of Peace Research 46(1): 79–97.

Lee, Chia-yi. forthcoming. Terrorism, Counterterrorism Aid, and Foreign Direct Investment. Foreign Policy Analysis.

Levy, Brian. 2014. Working with the Grain: Integrating Governance and Growth in Development Strategies. Oxford, UK: Oxford University Press.

Licht, Amanda. 2010. Coming into Money: The Impact of Foreign Aid on Leader Survival. Journal of Conflict Resolution 54(1): 58-87.

Light, M. 2014. Police reforms in the Republic of Georgia: the convergence of domestic and foreign policy in an anti-corruption drive. Policing and Society 24(3): 318–345.

Loayza, Norman, Jamele Rigolini, and Gonzalo Llorente. 2012. Do Middle Classes Bring Institutional Reforms? World Bank Policy Research Working Paper 6015. Washington DC: World Bank.

Lombardi, Domenico and Ngaire Woods. 2008. The Politics of Influence: An Analysis of IMF Surveillance. Review of International Political Economy 15 (5): 709–737.

Loud, Marlène Läubli and John Mayne. 2014. Enhancing Evaluation Use: Insights from Internal Evaluation Units. Sage Publishing.

Lu, Chunling, Catherine M Michaud, Emmanuela Gakidou, Kashif Khan, Christopher J. L. Murray. 2006. Effect of the Global Alliance for Vaccines and Immunisation on diphtheria, tetanus, and pertussis vaccine coverage: An independent assessment. Lancet 368: 1088–1095.

Luongo, Michael. 2012. A Rush to Do Business in the Middle East. July 16, 2012. Accessed online at http://www.nytimes.com/2012/07/17/business/global/after-the-arab-spring-a-rush-to-do-business-in-the-middle-east.html?pagewanted=all

March, James G. and Johan P. Olsen. 1998. The institutional dynamics of international political orders. International Organization 52: 943–969.

Marchesi, Silvia and Jonathon P. Thomas. 1999. IMF Conditionality as a Screening Device. Economic Journal 109: C111–25.

Martin, Facundo, Camilo Mondragón-Vélez and Laura Vila. 2014. Investing on Business Climate and Reform in Emerging Markets: An Equity Portfolio Performance Evaluation. Paper presented at the Doing Business: Past, Present and Future of Business Regulation Conference at Georgetown University's McDonough School of Business. Washington, DC. February 20-21, 2014. Accessed at http://www.doingbusiness.org/special-features/conference/~/media/GIAWB/Doing%20Business/Documents/Miscellaneous/Conference2014/S2-2.pdf

Mansfield, Edward D. and Jon C. Pevehouse. 2006. "Democratization and International Organizations." International Organization 60 (1): 137--167.

Martin, Lisa L. 2000. Democratic Commitments: Legislatures and International Cooperation. Princeton, N.J.: Princeton University Press.

McMillan, John and Pablo Zoido. 2004. How to Subvert Democracy: Montesinos in Peru. Journal of Economic Perspectives 18(4): 69-92.

Meessen, Bruno, Agnès Soucat, and Claude Sekabaraga. 2011. Performance-based financing: just a donor fad or a catalyst towards comprehensive health-care reform? Bulletin of the World Health Organization 89: 153-156.

Merchant-Vega, Nina, and Edmund Malesky. 2011. A Peek under the Engine Hood: The Methodology of Subnational Economic Governance Indices. Hague Journal on the Rule of Law 3(2): 186–219.

Millennium Challenge Corporation (MCC). 2005. Report on the Criteria and Methodology for Determining the Eligibility of Candidate Countries for Millennium Challenge Account Assistance in FY 2006. Washington D.C.: MCC.

Millennium Challenge Corporation (MCC). 2007. Report on the Criteria and Methodology for Determining the Eligibility of Candidate Countries for Millennium Challenge Account Assistance in FY 2008. Washington D.C.: MCC.

Mody, Ashoka, and Diego Saravia, 2003. Catalyzing Private Capital Flows: Do IMF- Supported Programs Work as Commitment Devices? IMF Working Paper. Washington DC: International Monetary Fund.

Molina-Gallart, Nuria. 2008. Outcome-based conditionality: Too good to be true? Brussels: EURODAD. http://www.eurodad.org/uploadedfiles/whats_new/reports/outcome_based_conditionality_final_a ugust_2010.pdf

Momani, Bessma. 2007. IMF Surveillance and America's Turkish Delight. Central European Review of International Affairs 27: 5-24.
Montinola, Gabriella R. 2010. When Does Aid Conditionality Work? Studies in Comparative International Development 45:358–382.
Moravcsik, Andrew. 2000. The Origins of Human Rights Regimes: Democratic Delegation in Postwar Europe. International Organization 54 (2): 217--252.
Morrison, K. M. 2009. Oil, nontax revenue, and the redistributional foundations of regime stability. International Organization 63 (Winter): 107–38.
Morse, Julia C. 2016. Pathways to Policy Change: Blacklists and Market Enforcement in the Regime to Combat Terrorist Financing. Paper Prepared for the Assessment Power in World Politics Conference, May 6-7 2016.
Murdie, Amanda and Dursun Peksen. 2015. Women's Rights INGO Shaming and the Government Respect for Women's Rights. The Review of International Organizations 10 (1): 1-22.
Muth, John F. 1961. Rational Expectations and the Theory of Price Movements. Econometrica 29 (3): 315-35.
Nelson, Stephen C. 2009. Endangered Species: Exploring the Political Survival of Economic Policymakers. Paper presented at the Cornell PSAC Seminar. April 2009.
Nelson, Stephen C. 2014. Playing Favorites: How Shared Beliefs Shape the IMF's Lending Decisions. International Organization 68: 297-328.
Neumayer, Eric. 2003b. The Pattern of Aid Giving: The Impact of Good Governance on Development Assistance. London, UK: Routledge.
Neumayer, Eric. 2005. Do International Human Rights Treaties Improve Respect for Human Rights? Journal of Conflict Resolution 49 (6): 925--953.
Nielsen, Richard. 2013. Rewarding Human Rights? Selective Aid Sanctions against Repressive States. International Studies Quarterly 57 (4): 791-803.
Noland, Marcus. 1997. Chasing Phantoms: The Political Economy of USTR. International Organization 51: 365-87.
Ölcer, Dilan. 2009. Extracting the Maximium from the EITI. OECD Working Paper No. 276. Paris: OECD.
Öhler, Hannes, Peter Nunnenkamp, and Axel Dreher. 2012. Does Conditionality Work? A Test for an Innovative US aid scheme. European Economic Review 56: 138-53.
Parks, Bradley C. 2014. Brokering Development Policy Change: The Parallel Pursuit of Millennium Challenge Account Resources and Reform. Ph.D. Dissertation. London, UK: London School of Economics and Political Science.
Parks, Bradley C. and Caroline Davis. 2017. When Do Governments Trade Domestic Reforms for External Rewards? Explaining Policy Responses to the Millennium Challenge Corporation's Eligibility Standards. Working Paper.
Parks, Bradley C. and Zachary J. Rice. 2013. Measuring the Policy Influence of the Millennium Challenge Corporation: A Survey-Based Approach. Working Paper. Williamsburg, VA: The College of William and Mary.
Parks, Bradley, Zachary Rice, and Samantha Custer. 2015. Marketplace of Ideas for Policy Change: Who do Developing World Leaders Listen to and Why? Williamsburg, VA: AidData and The College of William and Mary.
Patton, M.Q. 2008. Utilization focused evaluation. 4th Edition. Saint Paul, MN: Sage Publishing.
Perakis, Rita and William Savedoff. 2015. Does Results-Based Aid Change Anything? Pecuniary Interests, Attention, Accountability and Discretion in Four Case Studies. CGD Policy Paper 053. Washington DC: Center for Global Development.

Pevehouse, Jon C. 2002. With a Little Help From my Friends? Regional Organizations and the Consolidation of Democracy. American Journal of Political Science 4 6(3): 611-626.

Perrin, Burt. 2002. Implementing the vision: Addressing challenges to results-focused management and budgeting. Paris: OECD. http://www.seachangecop.org/sites/default/files/documents/2002%20Resultsfocused%20managing%20and%20budgeting.pdf

Perry, June Carter. 2008. "We have the Will and the Energy" - GOSL Makes Case for MCC. 01 December 2008. Accessed at http://www.cablegatesearch.net/cable.php?id=08FREETOWN560

Pham, J. Peter. 2009. Behind the Curve: Corrupt governments cash in on the Millennium Challenge Corporation's outdated metrics. Stanford Social Innovation Review Fall 2009.

Pitlik, Hans, Björn Frank, and Mathias Firchow. 2010. The Demand for Transparency: An Empirical Note. Review of International Organizations 5: 177-195.

Pop-Eleches, Grigore. 2009. From Economic Crisis to Reform: IMF Programs in Latin American and Eastern Europe. Princeton, NJ; Princeton University Press.

Pritchett, Lant, Michael Woolcock, and Matt Andrews. 2013. Looking Like a State: Techniques of Persistent Failure in State Capability for Implementation. The Journal of Development Studies 49 (1): 1-18.

Quiroz Flores, Alejandro, and Alastair Smith. 2013. Leader Survival and Natural Disasters. British Journal of Political Science 43 (4): 821-843

Radelet, Steve. 2006. "Pulling not Pushing Reforms: Delivering Aid through Challenge Grants." in The New Public Finance: Responding to Global Challenges, edited by Inge Kaul and Pedro Conceicao. NewYork and Oxford: Oxford University Press.

Reitmaier, Angela. 2014. Funding African Governance Institutions and Processes: Insights and Lessons. SAIIA Occasional Paper #117. Johannesburg, South Africa: SAIIA.

Rodrik, Dani. 1996. "Why is there Multilateral Lending?" In Annual World Bank Conference on Development Economics, edited by B.M. Arvin. Westport CT: Praeger.

Rodrik, Dani. 2007. One Economics, Many Recipes: Globalization, Institutions and Economic Growth. Princeton, NJ: Princeton University Press.

Russ-Eft, D. and H. Preskill. 2009. Evaluation in Organizations: A Systematic Approach to Enhancing Learning, Performance, and Change. (Second Edition). Cambridge, MA: Perseus.

Samuel, Boris. 2014. Economic Calculations, Instability and (In)formalization of the State in Mauritania, 2003--2011. Canadian Journal of Development Studies 31 (1): 77-96.

Sandefur, Justin and Amanda Glassman. 2015. "The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics." The Journal of Development Studies 51(2): 116-132.

Sartori, Anne E. 2005. Deterrence by Diplomacy. Princeton, NJ: Princeton University Press.

Savedoff. William. 2011. Incentive Proliferation? Making Sense of a New Wave of Development Programs. Center for Global Development Essay. Washington D.C.: Center for Global Development.

Soederberg, Susanne. 2004. American empire and 'excluded states': the millennium challenge account and the shift to pre-emptive development. Third World Quarterly 25 (2): 279-302.

Schimmelfennig, Frank and Ulrich Sedelmeier. 2004. Governance by conditionality: EU rule transfer to the candidate countries of Central and Eastern Europe. Journal of European Public Policy 11(4): 669–687.

Schueth, Samuel J. 2011. Assembling International Competitiveness: The Republic of Georgia, USAID, and the Doing Business Project. Economic Geography 37(1): 51-77.

Seyedsayamdost, Elham. Forthcoming. Millennium Development Goals: Impact on National Strategies and Spending. Development Policy Review.

Sharman, Jason. 2008. Power and Discourse in Policy Diffusion: Anti-Money Laundering in Developing States. International Studies Quarterly 635-56.

Sharman, Jason. 2009. The Bark is the Bite: International Organizations and Blacklisting. Review of International Political Economy 16: 573-596.

Sharman, Jason. 2011. Small is Beautiful: The Perils of Gigantism and the Pay-offs of Studying Small International Organisations. Presented at the AGORA Workshop on Institutional Diversity in Governance of the Global Economy, Brisbane, Australia, January 11-12, 2011.

Silva, Héctor. 2007. Cuidado con los $461 Millones! La Prensa Grafica. November 25, 2007.

Simmons, Beth A. 2000. The Legalization of International Monetary Affairs. International Organization 54: 573-602.

Sjöstedt, M. 2013. Aid effectiveness and the Paris Declaration: A mismatch between ownership and results-based management? Public Administration and Development 33: 143–155.

Smith, Alastair and James Raymond Vreeland. 2006. "The Survival of Political Leaders and IMF Programs." in Globalization and the Nation State: The Impact of the IMF and the World Bank, edited by G. Ranis, J. R. Vreeland and S. Kosack. New York: Routledge. Pp. 263- 289.

Srivastava, Vivek and Marco Larizza. 2013. Working with the Grain for Reforming the Public Service: A Live Example from Sierra Leone. International Review of Administrative Sciences 76 (3): 458–85.

Stokes, Susan. (ed.). 2001. Public Support for Market Reforms in New Democracies. Cambridge, UK: Cambridge University Press.

Stone, Randall W. Controlling Institutions: International Organizations and the Global Economy. Cambridge: Cambridge University Press.

Svensson, Jacob. 1998. Investment, Property Rights and Political Instability: Theory and Evidence. European Economic Review 42: 1317-1341.

Swedlund, Haley. 2013. The domestication of governance assessments: Evidence from the Rwandan Joint Governance Assessment. Conflict, Security & Development 13 (4): 449–470.

Tansey, O. 2007. Process tracing and elite interviewing: a case for non-probability sampling. PS: Political Science and Politics 40 (4):765-772.

Taylor, John B. 2004. Economic Freedom and Georgia's Rose Revolution. Remarks at the Caucasus Business School, Tbilisi, Georgia. November 22, 2004. http://www.stanford.edu/~johntayl/taylorspeeches/Economic%20Freedom%20and%20Georgia%20(22%20Nov%2004).doc

Thomas, M.A. 2010. What Do the Worldwide Governance Indicators Measure? European Journal of Development Research 22(1): 31–54.

Tingley, Dustin, anbd Barbara Walter. 2011. Reputation Building in International Relations: An Experimental Approach. International Organization 65: 343-365.

Tomz, Michael. 2007. Reputation and International Cooperation: Sovereign Debt Across Three Centuries. Princeton, NJ: Princeton University Press.

Trapnell, Stephanie E. 2011. Actionable Governance Indicators: Turning Measurement into Reform. Hague Journal on the Rule of Law 3 (22): 317-348.

Ubaldi, Barbara. 2013. "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives." OECD Working Papers on Public Governance, No. 22.

Unsworth, Sue. 2009. What's Politics Got to Do With It? Why Donors Find it So Hard to Come to Terms with Politics, and Why this Matters. Journal of International Development 21(6): 883-894.

UNDP and Global Integrity 2008. A User's Guide to Measuring Corruption. Oslo, Norway:

UNDP Oslo Governance Centre.
USAID Afghanistan. 2012. Fact Sheet: Investment Climate Reform Program. Kabul: USAID Afghanistan.
USAID. 2009. Georgia: Opened for Business. Georgia Business Climate Reform Final Report, Washington DC: USAID. Accessed at http://pdf.usaid.gov/pdf_docs/Pdacn591.pdf
Vreeland, James Raymond. 2003. The IMF and Economic Development. Cambridge, UK: Cambridge University Press.
Ward, V., House, A., and Hamer, S. 2009. Knowledge brokering: the missing link in the evidence to action chain? Evidence & Policy 5 (3): 267–279.
Wei, S.-J. 2000. How Taxing is Corruption on International Investors? The Review of Economics and Statistics 82(1): 1–11.
Weyland, Kurt Gerhard. 2005. Theories of Policy Diffusion: Lessons from Latin American Pension Reform. World Politics 57 (2): 262–95.
Wheatley, Jonathan. 2005. Georgia from National Awakening to Rose Revolution: Delayed Transition in the Former Soviet Union. Burlington, VT: Ashgate.
Wilde. 2011. "The Democratization of Governance Assessments." In Making the State Responsive: Experience with Democratic Governance Assessments, Edited by Göran Hydén and John Samuel. New York: UNDP.
World Bank. 2007. Celebrating Reform 2007: Doing Business Case Studies. Washington D.C.: World Bank.
World Bank. 2012. Fighting Corruption in Public Services: Chronicling Georgia's Reforms. Washington DC: World Bank.
World Bank. 2011. World Development Report 2012: Gender Equality and Development. http://siteresources.worldbank.org/INTWDR2012/Resources/7778105-1299699968583/7786210-1315936222006/Complete-Report.pdf
Wright, Joseph. 2008. To Invest or Insure? How Authoritarian Time Horizons Impact Foreign Aid Effectiveness. Comparative Political Studies 41(7): 971-1000.
Wroughton, Lesley. 2006. U.S. fund OKs aid for El Salvador, expands list. Reuters. 8 November 2006.