

Assessing International Organizations

Competition, Collaboration, and Politics of Funding

Ranjit Lall*

INTRODUCTION

During the past decade, states have initiated a striking trend in global governance: the development and promulgation of indicators of the performance of international organizations (IOs). These assessments possess all of the defining characteristics of global performance indicators (GPIs) outlined by Kelley and Simmons in the introduction to this volume. They are available to citizens of any country free of charge (*publicity*); they cover multiple IOs, including all those that regularly receive substantial funding from the government in question (*inclusiveness*); they rate IOs on a common numerical or ordinal scale (*comparativeness*); and they are typically updated on a consistent multiyear basis (*regularity*). Perhaps most importantly from an analytical perspective, the assessments are *purposive*: they were created to help policymakers make more efficient use of their multilateral funding in the context of budgetary pressures induced by the global economic crisis. In other words, they were conceived with the explicit intention of influencing financial resource flows to IOs.

A cursory examination of funding patterns since the emergence of the indicators, however, reveals an intriguing puzzle: only resource flows to *some* assessed IOs show signs of responsiveness to performance ratings. That is, for one subset of assessed IOs, there is a strong positive relationship between ratings and resource flows: high ratings have been met with a sizable increase in financial contributions, whereas low ratings have been followed by funding

* For valuable feedback on earlier drafts of this chapter, I am indebted to Daniel Drezner, Jeffrey Frieden, Hyeran Jo, Judith Kelley, Christopher Lucas, Walter Mattli, Beth Simmons, Anton Strezhnev, and participants in the 2016 Conference on Assessment Power at Harvard University and the Mini-conference on Global Assessment Power at the 2016 Annual Meeting of the American Political Science Association. Supplementary materials can be accessed at the following link: https://scholar.harvard.edu/files/ranjitlall/files/online_appendix_lall_ch10_pgpi.pdf.

cuts or freezes. For instance, while the strongly rated Office of the United Nations High Commissioner for Refugees (UNHCR) has seen its annual contributions rise by more than 50 percent since the release of its first set of ratings, the poorly rated Joint United Nations Programme on HIV/AIDS (UNAIDS) has seen them fall by almost one fifth. For another subset of assessed IOs, by contrast, there is no clear relationship between ratings and resource flows, with high-rated institutions seeing no financial “reward” and low-rated ones avoiding any “punishment.” Despite receiving similar ratings to UNHCR, for example, the European Development Fund (EDF) has suffered a decline in contributions comparable to that of UNAIDS. Conversely, despite receiving similar ratings to UNAIDS, the United Nations Environment Programme (UNEP) has enjoyed an increase in funding comparable to that of UNHCR.

These surprising differences pose a challenge to the few existing theories of the determinants of IO funding. Traditional analyses of the budgetary process in IOs draw attention to how procedural rules and practices constrain the ability of states to alter the size and composition of the budget, resulting in a high degree of stability in funding levels over time.¹ Yet as the above examples suggest, resource flows to many IOs have exhibited significant change since the release of the performance indicators. The more recent literature on the concept of delegation to IOs highlights the tendency of “state principals” to use the budget as an instrument for deterring and sanctioning opportunistic behavior by “IO agents,” implying that they will respond to performance indicators by increasing funding for IOs with strong ratings and reducing funding for IOs with weak ratings.² Once again, however, this prediction is not consistent with the pattern described earlier: only some high-rated institutions have received additional contributions, while only some low-rated ones have been subject to cuts. This pattern suggests the need for a more nuanced analysis of the factors affecting whether states and other donors are responsive to ratings. In other words, it calls for an answer to the question: *Under what conditions do IO performance indicators influence resource flows to these institutions?*

Analyzing ratings as a form of social knowledge that draws attention and gives meaning to institutional performance, I develop an argument that highlights how different aspects of the *relationship* between IOs and other actors within their policy space mediate the financial effects of such knowledge. Two aspects are particularly important. The first is the degree of competition IOs face from institutions performing similar functions to them. When competition is intense – which is most likely in policy spaces characterized by low barriers to entry and diseconomies of

¹ Ackrill and Kay 2006; Hoole, Handley, and Ostrom 1979; Hoole, Job, and Tucker 1976; Singer 1961.

² As one influential study of delegation states, “Agents that are perceived as succeeding in their missions are rewarded with larger budgets, allowing individuals to perform their jobs more easily or supervise larger staffs with compensatory benefits. Agents that are perceived as failing are punished with smaller budgets, and may even be eliminated entirely.” Hawkins et al. 2006, 30. Also see Nielson and Tierney 2003; Pollack 1997.

scale in the provision of goods – donors act on material and social incentives to respond to ratings because IOs have a large number of substitutes to which they can reallocate resources. When competition is limited, however, IOs have few or no substitutes, deterring donors from either sanctioning low-rated IOs or rewarding high-rated ones. The second aspect is the robustness of governance partnerships between IOs and non-state actors – both domestic and transnational – who possess complementary skills and resources and can thus enhance their capacity to achieve organizational objectives. I argue that robust partnerships render resource flows more responsive to ratings by incentivizing partners to assist high-rated IOs in mobilizing additional funds but – perhaps surprisingly – to weaken or exit partnerships with low-rated IOs, exacerbating the reputational damage suffered by such institutions and raising fears that they may perform even worse in the future. In short, the financial consequences of IO performance indicators are conditioned in critical ways by institutional relationships of competition and collaboration.

I test the argument using a mixed-methods approach. I begin by examining qualitative evidence on the assessments' financial effects from a selection of primary and secondary sources – most notably interviews with more than 80 IO officials and donor representatives – probing different behavioral as well as non-behavioral implications of the argument. I then conduct statistical tests of the argument based on an original dataset including all six sets of performance indicators issued thus far, which collectively cover 54 IOs spanning a wide range of issue areas. My methodological strategy involves analyzing – in both observational and quasi-experimental settings – the relationship between post-assessment changes in resource flows and multiplicative interaction terms between each set of indicators and the two conditioning variables highlighted by my argument. To gather information on institutional competition, I designed and implemented a survey of officials from all 54 assessed IOs. Data on governance partnerships were collected through a comprehensive mapping exercise covering all forms of substantive collaboration between IOs and non-state actors during five key stages of the international policymaking process.

By theorizing and empirically examining the financial consequences of IO performance indicators, the chapter contributes to our understanding of global assessment power in three ways. First, it draws attention to a substantively significant yet previously neglected area in which GPIs have emerged in recent years. In doing so, it provides a fitting counterpoint to the other contributions to this volume, most of which examine GPIs that evaluate states and are developed by IOs.³ Second, it presents the first systematic evidence that GPIs can influence resource allocation decisions by public actors, extending recent analyses in the economics and management literatures demonstrating their impact on the behavior of private investors.⁴ Such evidence, moreover,

³ See Bisbee et al.; Kelley, Simmons, and Doshi, and Morse, this volume.

⁴ Aaron, McMillan, and Cline 2012; Brammer, Brooks, and Pavelin 2009; Murguia and Lence 2015.

comes from an area in which the material stakes are high: the 54 assessed IOs have jointly received more than half a trillion dollars in contributions over the past decade. Third, and most significantly from a theoretical perspective, it sheds new light on *when* and *why* – and not just whether – GPIs matter in world politics by examining the conditions under which they exert social influence. My argument complements existing analyses of GPI scope conditions (in this volume as well as in fields such as anthropology and international law), which mostly focus on characteristics of GPIs and their creators, by highlighting how the broader relational context in which assessed entities operate mediates their effects.⁵ It thus enhances our analytical understanding of the mechanisms by which shared information and knowledge can become a source of power in the contemporary international system – and, equally important, the limits of such power.

The chapter also makes contributions to a number of other emerging areas of research in international relations. First, it complements recent work on institutional choice and organizational ecology in global governance by showing how institutional environments influence the effects of common information and knowledge as well as patterns of long-run institutional change.⁶ Second, and relatedly, it joins a growing number of studies in investigating the consequences of institutional competition at the global level, an issue that has traditionally received little attention in the international relations literature.⁷ Unlike previous scholarship, however, it examines the material as well as political ramifications of competition and gathers systematic data on the variable covering a diverse sample of institutions. Third, it extends recent research on the phenomenon of IO orchestration by drawing attention to the critical role of governance partnerships in conditioning the financial effects of IO performance indicators.⁸ Finally, as suggested above, it contributes to the burgeoning literature on delegation and principal-agent theory by showing – contrary to the conventional wisdom – that state principals only financially reward and sanction IO agents in response to performance information under certain relational conditions.⁹

IO PERFORMANCE ASSESSMENTS: OVERVIEW AND PUZZLE

To illustrate the puzzling variation in the relationship between IO performance indicators and resource flows, I begin with a brief overview of the assessments. As summarized in Table 10.1, since 2008 indicators have been

⁵ See Cooley and Snyder 2015; Davis, Merry, and Kingsbury 2015; Davis et al. 2012; Kelley and Simmons; Lee and Matanock, this volume. Note that Bisbee et al. argue that GPA effects are mediated by characteristics of assessed entities themselves (namely, democratic accountability and transparency).

⁶ Abbott, Green, and Keohane 2016; Jupille, Mattli, and Snidal 2013.

⁷ Abbott, Green, and Keohane 2016; Alter and Meunier 2009; Frey 2008; Lipsky 2015.

⁸ Abbott et al. 2015; Mattli and Seddon 2015.

⁹ See, for instance, Hawkins et al. 2006; Nielson and Tierney 2003; Pollack 1997.

TABLE 10.1. Summary of IO performance assessments

	Assessment						MOPAN	
	Australia	United Kingdom	Denmark	Netherlands	Sweden	Survey	Review	
Year(s)	2012	2011, 2016	2012	2011, 2013, 2015	2008–2011	2010–2014		
IOs	41	41	17	31	23	16		
Main data sources	Stakeholder consultations, written submissions; field visits; MOPAN and Paris Aid surveys; other GPIs, e.g. QuODA, Publish What You Fund ATI, COMPAS indicators		MOPAN surveys; ratings from British assessment	IO documents; internal/external evaluations; audits; feedback from overseas missions; MOPAN surveys	IO documents; feedback from overseas missions	Cross-national survey of public and private stakeholders	IO documents (analyzed by two consulting firms)	
Indicators	Delivery of results; contribution to international community's objectives; cost consciousness; financial management; transparency and accountability; strategic/performance management		One overall measure	Results control, strategy, focus on core mandate, effective governance, policy evaluation, human resources policy, financial management, anti-corruption policy	Internal and external effectiveness	Providing direction for results; corporate, country focus on results; thematic priorities; resource allocation; performance-oriented allocation, programming, financial accountability; using, presenting performance information; managing human resources; evaluating results; learning lessons		

		Assessment					
		Australia	United Kingdom	Denmark	Netherlands	Sweden	MOPAN
Year(s)	IOs	2012	2011, 2016	2012	2011, 2013, 2015	2008–2011	2010–2014
		41	41	17	31	23	16
Scale		Numerical (1–4)		Numerical (3,2–4,8)	Numerical (1–4)	Categorical (6 groups)	Numerical (1–6)
<i>r</i> within assessment		0.49	0.57	–	0.40	0.89	0.67
<i>r</i> across assessments		0.60	0.58	0.31	0.44	0.32	0.31
							0.36

Abbreviations: MOPAN = Multilateral Organisation Performance Assessment Network; QuODA = Quality of Official Development Assistance; ATI = Aid Transparency Index; COMPAS = Common Performance Assessment System.

produced by five states – Australia, Denmark, the Netherlands, Sweden, and the United Kingdom – and the Multilateral Organization Performance Assessment Network (MOPAN), a group comprising 17 of the largest donor countries (including all of the previous five) that seeks to enhance knowledge of IO effectiveness.¹⁰ Four of the six sets of indicators have been issued more than once, with the Australian and Danish assessments still awaiting their first update. The indicators are based on a mixture of new and existing data sources, including cross-national surveys, stakeholder consultations, feedback from diplomatic missions, and other GPIs such as the Quality of Official Development Assistance (QuODA) Assessment and the Publish What You Fund Aid Transparency Index (ATI).¹¹ As mentioned earlier, they collectively cover 54 IOs – a full list is provided in the chapter appendix – that span a diverse range of issue areas, including agriculture, development, the environment, human rights, public health, and trade. The Australian and British assessments have the widest organizational coverage (41 IOs each), while the MOPAN assessment has the narrowest (16 IOs).

International organizations are assessed on six distinct dimensions of institutional performance: (1) delivery of results; (2) cost and value consciousness; (3) resource management; (4) accountability and transparency; (5) strategic management; and (6) knowledge management.¹² The Australian, British, and Dutch assessments include between six and eight discrete numerical indicators of these dimensions ranging from 1 (“Weak”/“Unsatisfactory”) to 4 (“Very Strong”/“Strong”). The MOPAN assessment contains 14 numerical indicators measuring the third, fifth, and sixth dimensions. The IOs are assigned two scores from 1 to 6 on each indicator, one based on a multicountry stakeholder survey (a continuous scale) and other based on a review of organizational documents by two consulting firms (a discrete scale). The Swedish assessment contains two categorical indicators, “internal effectiveness” and “external effectiveness,” on which IOs are ranked in six classes spanning from “Very Low” to “Very High.”¹³ Finally, the Danish assessment contains just one overall performance measure, a continuous numerical scale from 3.2 to 4.8.

¹⁰ The documents comprising each set of ratings can be found at the following links (all accessed February 21, 2015): <http://dfat.gov.au/about-us/publications/Pages/australian-multilateral-assessment-ama-full-report.aspx> (Australia); <http://www.ft.dk/samling/20111/almindel/uru/bilag/245/1153552.pdf> (Denmark); <http://www.mopanonline.org/Assessments/> (MOPAN); <https://www.rijksoverheid.nl/zoeken?trfwoord=scorecard> (Netherlands); <http://www.government.se/search/?query=%22swedish+assessment%22> (Sweden); <https://www.gov.uk/government/collections/multilateral-aid-review> (United Kingdom).

¹¹ On the ATI, see Honig and Weaver this volume.

¹² All six correspond to one of the three dimensions of institutional performance traditionally emphasized in the organizational theory and public administration literatures on the concept: (1) the achievement of stated objectives; (2) cost-effectiveness; and (3) responsiveness to a wide range of stakeholders. See Cameron 1978; Etzioni 1964; and Price 1972. Some assessments include additional indicators that do not correspond to any of these dimensions and are thus excluded from consideration.

¹³ I convert these categories into a 1–6 numerical scale in the subsequent analyses.

Within the five assessments that contain multiple performance indicators, there is a strong positive correlation among ratings. That is, IOs with higher (or lower) ratings on some dimensions of performance also tend to have higher (or lower) ratings on others. As shown in Table 10.1, the mean correlation among indicator scores from each wave of a given assessment a between its initial release (year t_a) and 2015 is $r = 0.53$ (more than four-fifths of the 532 individual coefficients are positive and statistically significant at the 10 percent level). Interestingly, ratings are also strongly and positively correlated *across* assessments. The mean correlation between averages of each set of indicator scores in the period from year t_a to 2015 is $r = 0.42$ (17 of the 21 coefficients are positive and significant). In other words, there is a high degree of consensus among the assessments about which IOs are performing well and which are performing poorly.

Figure 10.1 provides a graphical overview of the relationship between performance ratings and over-time changes in resource flows to assessed IOs.

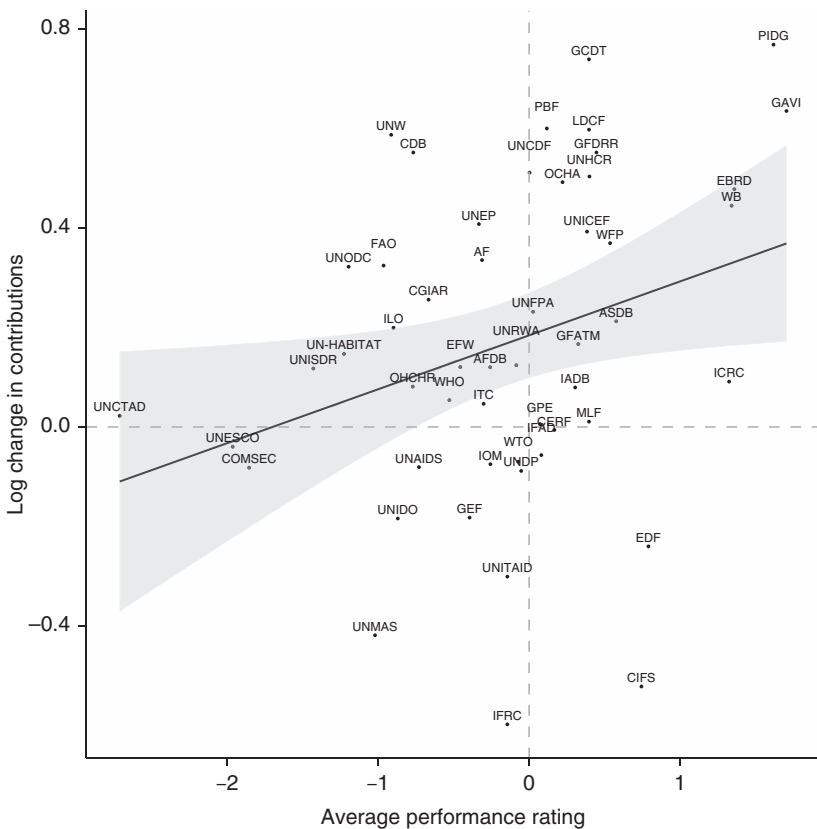


FIGURE 10.1. Relationship between performance ratings and changes in contributions.

The x -axis measures the standardized mean of an IO's indicator scores from all assessments averaged over the period from the release of its first set of ratings (year u_{IO}) to 2015. The y -axis measures the log ratio of an IO's average annual contributions since financial year (FY) u_{IO} to its contributions in the previous FY $u_{IO} - 1$:

$$y = \ln \left(\frac{\sum_{i=0}^n \text{CONTRIBUTIONS}_{u_{IO}+i}}{n \cdot \text{CONTRIBUTIONS}_{u_{IO}-1}} \right)$$

Data on contributions, which were collected from IO financial statements and annual reports (acquired online and in some cases through personal communications with officials), are all converted into millions of United States dollars at contemporary exchange rates and adjusted for inflation using the Bureau of Labor Statistics' Consumer Price Index (with 2003 as the base year).

The scatterplot reveals striking variation in post-assessment funding trends among both high- and low-rated IOs. For approximately half of the sample, trends are consistent with a straightforward "GPA influence hypothesis": IOs with higher ratings have received proportionally larger increases in contributions since FY u_{IO} (roughly the lower-left and upper-right quadrants of the graph). For the remaining institutions, however, trends provide little support for the hypothesis: IOs with below-average ratings have received disproportionately large increases in funding (upper-left quadrant), while IOs with above-average ratings have seen either disproportionately small increases or absolute decreases (lower-right quadrant). The upshot of these differences is that the overall relationship between x and y is positive but relatively weak, with only 16 of the 54 IOs falling inside the shaded 95 percent confidence interval around the regression line.

Figures 10.2 and 10.3 provide a micro-level view of such variation by displaying time-series data on resource flows to 12 individual IOs.¹⁴ For the six institutions in Figure 10.2, funding patterns are consistent with the GPA influence hypothesis: the first three (reading from left to right) have received high ratings and enjoyed robust growth in contributions since FY u_{IO} ; the last three have received low ratings and experienced weak or negative growth. Visual analysis of the timing of these changes provides *prima facie* evidence that they were a direct response to the ratings rather than a product of broader (exogenous) shifts in funding. The six institutions in Figure 10.3 display the opposite patterns: the first three have received high ratings, but subsequently suffered a

¹⁴ I return to a number of these institutions in the fourth section, providing qualitative evidence on the factors shaping their funding trends.

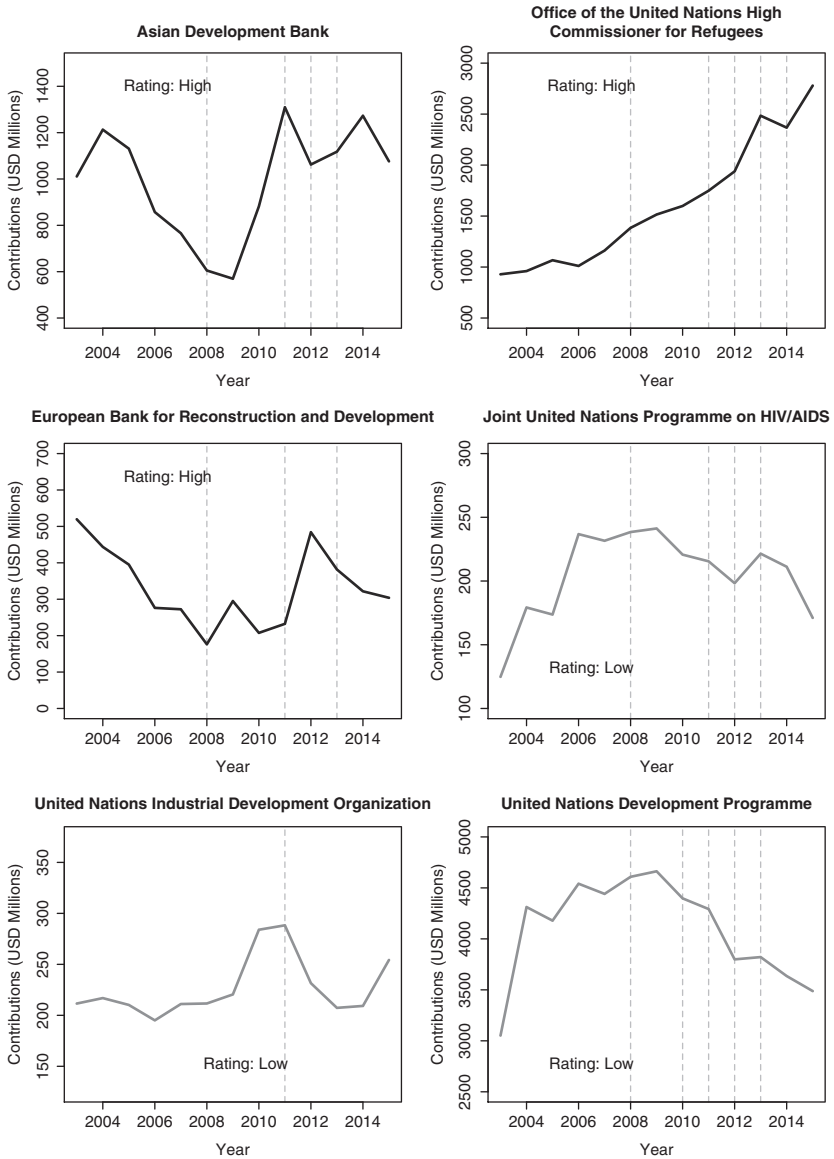


FIGURE 10.2. Funding trends consistent with indicator influence.

Note: The dotted lines indicate the years in which an IO has been assessed.

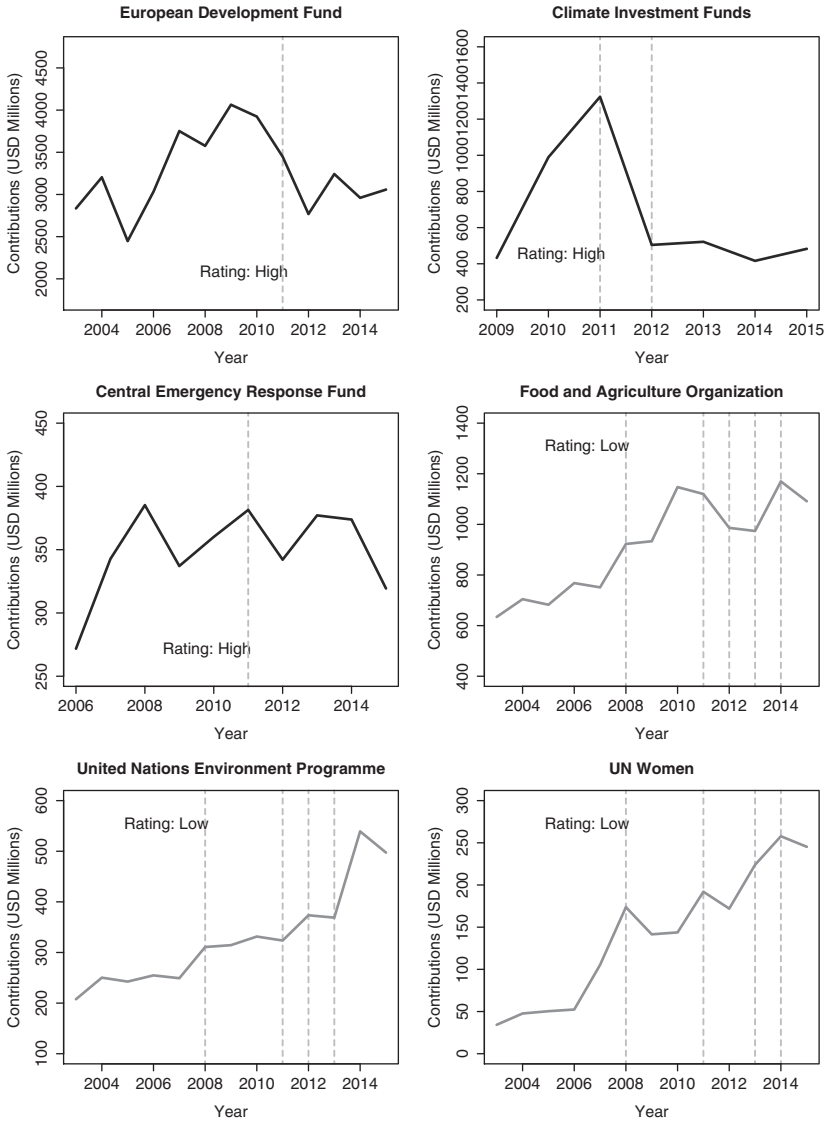


FIGURE 10.3. Funding trends inconsistent with indicator influence.

Note: The dotted lines indicate the years in which an IO has been assessed.

stagnation or decline in contributions; the last three have received low ratings, but seen strong growth. Unlike before, visual analysis of these trends suggests that they were largely unaffected by ratings. In short, in line with the evidence presented in Figure 10.1, only in a subset of assessed IOs do resource flows appear to have been responsive to performance indicators.

A RELATIONAL THEORY OF ASSESSMENT POWER

The IO performance indicators constitute a form of social knowledge produced by a set of respected actors whose authority stems from their institutional expertise, bureaucratic capacity, and central position in the “network” of multilateral donors. By defining and setting standards for institutional performance, they highlight, frame, and instill meaning in the behavior and effects of IOs. They thus not only increase the salience of institutional performance as a metric by which IOs are judged but also clarify and crystallize shared understandings of the concept. In doing so, they alter the material and normative calculus by which donors allocate resources to IOs, incentivizing them to place greater weight on performance considerations as a means of both maximizing the “return” on their contributions and enhancing their status and reputation in the donor community.¹⁵ Other things equal, therefore, the creation of IO performance indicators should bring about an increase in funding for high-rated institutions and a reduction for low-rated ones.

Yet other things are not always equal. IOs exist not in a vacuum but in a distinct transnational policy space populated by diverse actors from different institutional levels (subnational, national, and supranational) and sites of authority (public, private, and hybrid). What distinguishes these actors as inhabitants of a common policy space is that they seek to shape substantive outcomes in the same issue area, whether by performing, supporting, or influencing the exercise of governance functions. The central claim of my argument is that the financial consequences of IO performance indicators are contingent upon the *relationship* between these institutions and other actors within their policy space. Only under particular relational conditions do ratings affect resource flows to IOs through the three pathways of GPA influence delineated by Kelley and Simmons, that is, domestic politics, elite response, and transnational mobilization. I highlight two such conditions, the first mediating the elite pathway and the second mediating the domestic and transnational pathways: (1) IOs are subject to a high level of institutional competition; and (2) IOs possess robust governance partnerships with non-state actors.

Institutional Competition

The degree of institutional competition faced by IOs is a function of the number of institutions within their policy space that exercise similar governance functions to them. In most instances, these institutions are themselves IOs. While other types of institutions – such as treaties, agreements, and laws – often address similar policy problems to IOs, their lack of a permanent secretariat prevents them from performing many of the latter’s distinctive

¹⁵ In this way, reputational concerns can affect how actors linked to assessed entities – and not just assessed entities themselves – respond to GPIs.

governance functions, such as gathering and providing information, resolving disputes, allocating pooled resources, identifying emerging transnational problems, and formulating rules and standards.¹⁶

The political implications of competition among IOs have received surprisingly little attention from scholars of such institutions, a possible reflection of the traditional dominance of functionalist approaches to analyzing international cooperation, which imply that intense competition is unlikely to arise because overlap in governance functions is inefficient as well as redundant (since only one institution is needed to perform a given function).¹⁷ In reality, as recent studies have pointed out, there is substantial *variation* in the degree of competition experienced by IOs.¹⁸ Some policy spaces are indeed sparsely populated with functionally similar IOs, ensuring that each enjoys a monopoly or quasi-monopoly over governance functions. In other policy spaces, however, IOs share authority with a sizable population of competitors, each of which makes only a marginal contribution to the aggregate supply of institutional goods.

These differences are primarily a function of two issue-specific characteristics. The first is the presence of barriers to entry into the policy space, i.e. costs incurred in the exercise of governance functions that prevent IOs from establishing themselves as “suppliers.” A major barrier to entry in the mostly low-politics issue areas covered by IO performance indicators is the need for technical knowledge in the policymaking process.¹⁹ Such expertise is costly to acquire because it requires IOs to recruit staff with specialized qualifications – which may be in short supply and command a wage premium – and often to provide them with years of additional training and practical experience related to their specific tasks. The second characteristic is the existence of economies of scale in the provision of governance functions, i.e. declines in the average cost of provision as output increases. Perhaps the most common source of scale economies in low-politics domains is the occurrence of “network effects,” whereby the marginal utility of participating in an IO’s activities rises with the number of participants. This is because many IOs in such domains perform standard-setting functions, which give rise to network effects because each adoption of a given standard yields coordination benefits for all previous adopters.

How do differences in these characteristics across policy spaces mediate the financial effects of IO performance indicators? In policy spaces characterized by high barriers to entry or economies of scale, IOs have few or no close substitutes to which donors can reallocate resources. Given the high transaction

¹⁶ Abbott and Snidal 1998; and Pevehouse, Nordstrom, and Warnke 2004.

¹⁷ It may also reflect the field’s focus on a relatively small number of IOs that *do* face little competition, such as the International Monetary Fund (IMF) and the World Trade Organization (WTO).

¹⁸ Abbott, Green, and Keohane 2016; Frey 2008; Jupille, Mattli, and Snidal 2013; Lipsy 2015.

¹⁹ Lipsy 2015.

costs and uncertain distributional consequences of creating new institutions, donors are thus likely to refrain from imposing major financial sanctions on IOs receiving low performance ratings for fear of jeopardizing the supply of institutional goods in the policy space.²⁰ That is, material and social incentives to withdraw support from ineffective IOs are likely to be trumped by concerns about squandering *all* the benefits of institutionalized cooperation, blocking the elite response pathway of GPA influence. Nor, conversely, will donors have incentives to provide increased contributions to IOs receiving high ratings: since they cannot credibly threaten to punish such institutions for suboptimal performance in the future, rewards are unlikely to have their desired effect of incentivizing continued effectiveness.

In policy spaces characterized by low barriers to entry or diseconomies of scale, by contrast, IOs have a sizable pool of close substitutes. In accordance with the elite response pathway, therefore, donors can sanction low-rated IOs without fear of compromising the supply of institutional goods in the policy space. In addition, they have incentives to provide additional contributions to high-rated IOs: because the threat of sanctioning poor performance is now credible, rewards *are* likely to have their intended effect of incentivizing such institutions to continue performing effectively. In sum, I hypothesize that only a high degree of institutional competition will result in a strong positive relationship between performance indicators and resource flows.

Governance Partnerships

Governance partnerships are relational arrangements between IOs and actors above and below the state – including non-governmental organizations (NGOs), businesses, transgovernmental networks, and other IOs – involving voluntary and sustained collaboration in the exercise of governance functions.²¹ Such arrangements are based on a convergence of both goals and interests. IOs often lack the informational, epistemic, administrative, and logistical resources to successfully fulfill their mandates – a consequence of the ambitious nature of such mandates as well as the limited support they receive from their members – and enlist partners as a means of addressing these capacity deficits. Partners are willing to provide assistance because they share organizational objectives, perceive potential synergies in combining their capabilities with those of IOs, and derive considerable non-material benefits from collaboration, including access to contacts and networks, normative guidance, and the legitimacy that comes with endorsement by a multilateral institution.²²

²⁰ Jupille, Mattli, and Snidal 2013.

²¹ Closely related concepts include “joint governance,” “multi-stakeholder partnerships,” and “orchestration.” I follow Mattli and Seddon 2015 in using “governance partnerships” as an umbrella term for these various arrangements.

²² See Abbott and Snidal 1998; and Abbott et al. 2015.

While many IOs have established governance partnerships in recent years, there is substantial variation in the *robustness* of such arrangements. I define robustness as an increasing function of three factors. The first is the depth of cooperation between IOs and partners. Upon closer inspection, many partnerships turn out to be purely symbolic arrangements formed in order to satisfy external pressures for stakeholder engagement. I only consider partnerships to be robust if they involve substantive collaboration at one of the five principal stages of the IO policymaking process: agenda setting, formulation, monitoring, implementation, and enforcement. The second factor is the degree of alignment between the policy preferences of IOs and partners. Although partners share the broad objectives of IOs, their views about the specific policies needed to advance such objectives often conflict with those of IO officials. Misalignment is particularly common when there are differences in the geographical scope and range of issue areas covered by partners and IOs. The third factor is the level of complementarity between the capabilities of IO and partners. Naturally, partners with ample (material and non-material) resources are better able to remedy IO capacity deficits. Yet since the size of such deficits depends on an IO's own level of resources, complementarity increases not with the *absolute* level of partner resources but with the *ratio* of partner to IO resources.

Robust governance partnerships create incentives for non-state actors to behave in ways that enhance the sensitivity of resource flows to IO performance indicators. When an IO receives high ratings, the logic is straightforward: due to their close operational ties with the institution, support for its policies, and sizable resource base, partners are both willing and able to assist it in mobilizing additional contributions. Such assistance can take several forms, including lobbying governments and other donors at the domestic level; publicizing and disseminating information about the assessments; identifying and targeting potential new donors; and increasing their own contributions to the IO.²³ When robustness is low, partners are unlikely to invest in these mobilization strategies for want of both incentives (they have weak operational ties with the IO and non-aligned policy preferences) and means (they possess limited resources).²⁴ That is, the domestic and transnational pathways of GPA influence are unlikely to operate as specified.

When an IO receives low ratings, the effects of variation in partnership robustness are less clear. It may appear that when robustness is high, partners will seek to protect the IO from financial sanctions using the mobilization strategies mentioned above, mitigating the ratings' impact on resource flows. I argue, however, that they are more likely to respond in ways that *increase* the size of such sanctions. This is because the gains they derive from continued

²³ For examples of these strategies, see Broz and Hawes 2006; Busby 2007; and Lavelle 2011.

²⁴ Similarly, non-state actors who are not part of governance partnerships will often lack a sufficiently large stake in the IO's success to be willing to incur these costs.

collaboration will tend to be outweighed by (1) the reputational costs of association with a low-rated IO (which can include a reduction in their own funding) and (2) the opportunity costs of foregoing collaboration with more effective IOs (which can include reputational gains and an enhanced capacity to achieve their objectives). Partners thus have incentives to scale down or withdraw their support for the IO (including on the financial front) and establish ties with higher-rated institutions. Such actions, in turn, exacerbate the reputational damage suffered by the IO and raise concerns that it may exhibit even worse performance in the future. The upshot is likely to be a further intensification of sanctions. When partnerships are weak, this sequence of events is unlikely to transpire: since cooperation involves no meaningful exchange of resources or services, partners neither incur major (reputational or opportunity) costs from maintaining the relationship nor inflict major (reputational or operational) damage on the IO by withdrawing their support. In sum, my first hypothesis is that only a high degree of institutional competition will result in a strong positive relationship between performance indicators and resource flows.²⁵

QUALITATIVE EVIDENCE

How much support is there for these hypotheses? I begin my empirical investigation by examining a range of primary and secondary qualitative sources on the financial effects of IO performance indicators, including policy statements and budgetary documents from donor governments, media reports, and information from more than 80 interviews with IO officials and donor representatives conducted between 2012 and 2016.²⁶ In addition to providing a preliminary test of the hypotheses, this examination sheds light on the argument's posited causal mechanisms – both behavioral and non-behavioral – and offers concrete illustrations of how such mechanisms have shaped funding trends in individual institutions (including several of those featured in Figures 10.2 and 10.3).

The most direct evidence that IO performance indicators have influenced resource flows comes from the states that developed them. All five governments

²⁵ Note that both hypotheses imply a weak relationship between the conditioning variable and performance indicators. Indeed, there are obvious reasons why competition and partnerships are likely to have mixed effects on ratings. While competition puts pressure on IOs to perform well to secure funding, for instance, it can also lead to institutional overlap and “crowding out” that undermines effectiveness. Similarly, while robust partnerships may enhance the ability of IOs to perform governance functions, they can also dilute the influence of member states and thus facilitate shirking and slippage.

²⁶ The interviews were conducted in four waves: the first in Washington, DC, in spring 2012; the second in Geneva in summer 2012; the third in London in summer 2014; and the fourth in Rome in winter 2015. Additional interviews were conducted via telephone throughout the period.

that have conducted individual assessments have explicitly indicated that their findings have informed subsequent multilateral funding decisions.²⁷ In addition, a recent survey of MOPAN's 17 member nations – a group that accounts for 95 percent of the world's multilateral development assistance – revealed that all but two have used its evaluations to “decide on funding allocations about multilateral organizations.”²⁸ The only public time-series data on such allocations, which come from the British government, suggest that the assessments' impact has been significant.²⁹ As a direct result of its first set of ratings in 2011, the government eliminated its assessed funding for four low-rated IOs – the International Labour Organization (ILO), the UN Human Settlements Programme (UN-Habitat), the UN International Strategy for Disaster Reduction (UNISDR), and the UN Industrial Development Organization (UNIDO) – and more than doubled contributions to 11 high-rated IOs. Media reports suggest that a recent update of the ratings, released in December 2016, will result in a further £180m of cuts to poorly rated UN agencies – most notably the UN Educational, Scientific, and Cultural Organization (UNESCO) – causing them to lose almost all of their assessed British contributions.³⁰

Interviews with IO officials and representatives of donor states provide further evidence of such influence. Several of the latter described “triangulating” between different assessments in determining allocations – a feat rendered easier by the high correlation between them – with some even referring to them as the “single most salient” factor in the decision-making process. Interestingly, in line with the argument's emphasis on framing and meaning making, some interviewees also noted that the assessments have altered the official discourse around multilateral funding. According to an employee of the Swiss Agency for Development and Cooperation (SDC),

“The evaluations have changed how governments talk about and justify supporting IOs, causing them to frame funding decisions in terms of ‘efficiency’ and ‘value for money’ rather than the protection of national interests. Since the ratings effectively

²⁷ Ministry for Foreign Affairs (Sweden) 2011; Department for International Development (United Kingdom) 2013; Department of Foreign Affairs and Trade (Australia) 2013; Ministerie van Buitenlandse Zaken (Netherlands) 2013; and Ministry for Foreign Affairs (Denmark) 2013.

²⁸ Multilateral Organisation Performance Assessment Network 2015, 19.

²⁹ Statistics on International Development, Department for International Development (United Kingdom). Available at <https://www.gov.uk/government/organisations/department-for-international-development/about/statistics>, accessed December 14, 2016. Where possible, I supplement this source with data from the government's Development Tracker database (<https://devtracker.dfid.gov.uk/>, accessed December 14, 2015) and IO annual reports and financial statements. All figures are converted into millions of British pounds at contemporary exchange rates and adjusted for inflation (using 2000 as the base year).

³⁰ “‘Wasteful’ UNESCO targeted as Britain makes £180m funding cuts.” *The Sunday Times*, April 17, 2016, 20.

quantify and compare these characteristics across IOs, they have created a competitive dynamic among donors whereby supporting weakly rated IOs can lead to reputational damage.”³¹

The IO staff similarly attested to the assessments’ impact on funding decisions. Division heads in both UNAIDS and the United Nations Development Programme (UNDP), for instance, cited the combination of low ratings and intense competition from similar IOs as the primary reason for their institutions’ recent decline in contributions.³² The UNAIDS official also highlighted the ratings’ non-behavioral effects – effects strikingly parallel to those mentioned above – observing that “being directly compared with peers” had made UNAIDS’s “entire organizational culture more efficiency- and results-oriented” and heightened “internal awareness about our relative status and reputation.”³³

Is there other evidence that the financial effects of the assessments have been mediated by institutional competition and governance partnerships? Consistent with the evidence presented in the second section, not all IOs that received low ratings in the first British assessment have been financially sanctioned by the country, while not all IOs that received high ratings have been rewarded. A closer look at the “anomalous” cases suggests that resource flows have been less responsive to ratings when IOs are subject to limited competition and possess weak partnerships. Of the eight low-rated IOs placed in the lowest summary category “poor value for money,” for example, the only two that have avoided funding cuts are the UN Entity for Gender Equality and the Empowerment of Women (UN Women), the sole IGO (intergovernmental organization) with a mandate to promote women’s rights; and the Food and Agriculture Organization (FAO), which, in addition to being the only IGO responsible for monitoring food security and setting agricultural standards at the global level, has a notoriously poor reputation for collaborating with non-state actors.³⁴ Similarly, of the nine IOs in the highest “very good value for money” category, the only one that has failed to receive additional contributions is the EDF, a poverty reduction fund known to possess weak partnerships due to its distinctive “joint ownership” governance model, which allows recipient country governments – but not non-state actors – to play a direct role in designing and implementing development projects.³⁵

³¹ Author interview with employee of Swiss Agency for Development and Cooperation, June 9, 2012, Geneva.

³² Author interview with UNAIDS division director, June 12, 2012, Geneva; and author telephone interview with UNDP division director, May 21, 2018.

³³ Author interview with UNAIDS division director, June 12, 2012, Geneva.

³⁴ The World Food Programme (WFP) also operates in the realm of global food security, but performs the very different function of delivering emergency food aid on the ground. On the FAO’s weak partnerships, see Food and Agriculture Organization 2007.

³⁵ See Gavas 2012.

Interviews provide further evidence of the conditioning effects of competition and partnerships. Donor representatives frequently expressed concerns that sanctioning low-rated IOs with a dearth of close substitutes could jeopardize key global public goods and thus undermine foreign policy goals. As a senior bureaucrat in Italy's Ministry of Foreign Affairs explained, "While [ratings] do guide our funding decisions, it is not always in our interest to follow them. For instance, if we stop financing UNEP because it is poorly rated, who will lead the global response to global warming?"³⁶ Such concerns were also recognized by low-rated IOs themselves, with one UNEP official even suggesting that the agency had been "saved from life-threatening funding cuts" by its "unique niche in coordinating national efforts to address climate change."³⁷ A number of officials in high-rated IOs, meanwhile, highlighted how low levels of competition had weakened incentives for donors to reward them. One economist in the Multilateral Fund (MLF) for the Implementation of the Montreal Protocol, for instance, complained that the institution's strong ratings had not led to increased funding because "we're the only source of funding for mitigating ozone depletion, which makes it difficult for states to pull the plug if we perform badly in the future – and, as economists know, rewards don't work without a credible threat of sanctions."³⁸

Interviewees also drew attention to the key role of partnerships in shaping how donors respond to ratings. A recurring theme was the importance of robust partnerships – particularly those characterized by high levels of complementarity – in providing high-rated IOs with the political and organizational support necessary to mobilize additional funds. The following view, expressed by a financial officer in UNHCR, was typical:

"We've received consistently high scores in the evaluations, but wouldn't have received such a large increase in funding if it hadn't been for our major NGO partners, such as the International Rescue Committee, Save the Children, and the Scandinavian Refugee Councils... They've been incredibly effective in using their campaigning infrastructure to raise public awareness about the ratings and their political contacts to lobby wealthy donor governments – in particular the US – for increased contributions."³⁹

Government officials also acknowledged the influence of partners in their decision to reward high-rated IOs. One employee of the United States Agency for International Development (USAID), for instance, noted that its near threefold increase in annual contributions to UNHCR since 2009 is "in part the result of an aggressive ratings-focused lobbying drive by the agency's most well-resourced civil society partners."⁴⁰ Staff in low-rated IOs, by contrast, lamented

³⁶ Author interview with employee of Italy's Ministry of Foreign Affairs and International Cooperation, January 23, 2015, Rome.

³⁷ Author telephone interview with UNEP program officer, December 2, 2013.

³⁸ Author interview with MLF staff economist, July 14, 2018, Washington, DC.

³⁹ Author interview with UNHCR financial officer, June 6, 2012, Geneva.

⁴⁰ Author interview with employee of USAID, May 8, 2012, Washington, DC.

the unexpected tendency of robust partnerships to exacerbate the financial damage caused by the assessments. In the words of a partnerships coordinator in the Commonwealth, “Instead of using their clout with donors to protect us against funding cuts [resulting from low ratings], many of our most important civil society partners have weakened or severed ties with us, causing even greater alarm among donors. Unfortunately, the result has been yet deeper cuts.”⁴¹

STATISTICAL ANALYSIS

In this section, I subject the argument to a series of observational and quasi-experimental statistical tests based on a new panel dataset, part of which was introduced in the second section. This analysis complements the qualitative examination both by providing systematic cross-organizational data on the dependent and explanatory variables and by showing that the earlier findings are generalizable to a larger and more representative population of IOs. I begin by describing the contents of the dataset, before outlining my methodological strategy and presenting the results.

The main dependent variable $\Delta\text{CONTRIBUTIONS}_a$, which is similar to y in Figure 10.1, is the log ratio of an IO’s contributions in each FY since the release of a given assessment a ($\text{FY}_{v_{IO}}$) to its contributions in the FY preceding this date ($\text{FY}_{v_{IO}-1}$):

$$\Delta\text{CONTRIBUTIONS}_a = \ln \left(\frac{\text{CONTRIBUTIONS}_{v_{IO}+i}}{\text{CONTRIBUTIONS}_{v_{IO}-1}} \right)$$

RATING_a is the standardized mean of an IO’s indicator scores from the most recent wave of an assessment a in $\text{FY}_{v_{IO}+i}$. To maximize the sample size and capture the possibility that donors are “triangulating” between different sets of ratings, I supplement this measure with an aggregate index ($\text{RATING}_{\text{avg}}$) that averages its value across all assessments (similarly to x in Figure 10.1). As discussed below, in separate analyses I also disaggregate indicator scores within each assessment to examine whether resource flows are more responsive to information about some dimensions of performance than others.

Institutional competition is measured using responses to an original online survey of IO head officials conducted between September 2013 and January 2015.⁴² Participants were asked the following question for a selection of recent years:

⁴¹ Author interview with Commonwealth Secretariat partnerships coordinator, July 6, 2014, London.

⁴² The survey, which was implemented using the Qualtrics Survey Software, was sent to participants via an emailed link. In most cases, 2–3 reminder messages were sent before the response was submitted. In five cases, participants preferred to provide their responses either verbally (during a telephone interview) or in writing. To check the reliability of responses, in 20 percent of cases the survey was sent to another senior official (usually a division or department head). In no instances were there discrepancies between the two sets of answers, indicating a high degree of reliability.

“How many international organizations perform a similar function to your organization and thus might be seen to compete with it?” Five response options were provided: (1) “Zero”; (2) “Between 1 and 5”; (3) “Between 5 and 10”; (4) “Between 10 and 20”; and (5) “More than 20.” COMPETITION is constructed by converting responses for the year (FY_{v_{IO} - 1}): into a five-point scale ranging from 0 (corresponding to option 1) to 4 (corresponding to option 5).⁴³ I later experiment with alternative (non-survey-based) measures, some of which capture the issue-specific determinants of competition discussed in the third section.

PARTNERSHIPS is an additive index of four indicators measuring the three dimensions of governance partnership robustness outlined earlier. All indicators are measured as of FY_{v_{IO} - 1} using information from IO websites, most of which have a section devoted specifically to partnerships.⁴⁴ The first indicator measures the depth of cooperation between IOs and partners. IOs receive a score between 0 and 1 reflecting the proportion of partnerships that involve substantive collaboration at the agenda-setting, formulation, monitoring, implementation, or enforcement stage of the policymaking process (as opposed to a purely symbolic affiliation). The second and third indicators measure policy preference alignment. IOs are assigned a score from 0 to 1 based on the proportion of partners that share their geographical scope (second indicator) and range of issue areas (third indicator).⁴⁵ The fourth indicator measures complementarity in capabilities, taking a value of 1 if the ratio of combined partner-to-IO expenditures exceeds 10 and 0 otherwise. IOs that have no listed partnerships are assigned a score of 0 on each indicator.

I control for four financial variables that are likely to affect how resource flows respond to ratings. TREND_a is the average log ratio of an IO’s contributions in FY_{v_{IO} - j} to its contributions in the previous FY_{v_{IO} - j - 1} during the five-year period before the release of an assessment *a*:

$$\text{TREND}_a = \frac{\sum_{k=0}^5 \ln \left(\frac{\text{CONTRIBUTIONS}_{v_{IO}-j}}{\text{CONTRIBUTIONS}_{v_{IO}-j-1}} \right)}{5}$$

CONTRIBUTIONS is an IO’s log contributions in FY_{u_{IO} - 1}. VOLUNTARY is the proportion of an IO’s contributions in FY_{u_{IO} - 1} that are voluntary rather than assessed (i.e. not paid as a condition of IO membership). INDEPENDENT is the proportion of an IO’s total income in FY_{u_{IO} - 1} that is composed of independently earned revenue – for instance, from investments, interest charges, and fees for services rendered – as opposed to contributions.

⁴³ Responses varied little across years, perhaps reflecting that the fact that competition is primarily a function of issue-specific characteristics that tend to change slowly over time.

⁴⁴ To access older versions of these websites, I used the Internet Archive’s Wayback Machine. Available at <https://archive.org/web/>, accessed February 21, 2016.

⁴⁵ I distinguish between three levels of geographical focus – global, regional, and national – and 25 issue areas based on a list compiled by Hooghe et al. 2017.

Baseline Model and Results

Given the conditional nature of the argument, I estimate a linear multiplicative interaction model, clustering robust standard errors by IO to address potential cross-sectional dependence. In light of the relatively small sample size (due to the recent emergence of the assessments), I begin by testing the two hypotheses separately:

$$\begin{aligned} \Delta \text{CONTRIBUTIONS}_a = & \beta_0 + \beta_1 \text{RATING}_a + \beta_2 \text{COMPETITION} \\ & + \beta_3 \text{RATING}_a \times \text{COMPETITION} + \beta_4 \text{TREND}_a + \beta_5 \text{CONTRIBUTIONS} \\ & + \beta_6 \text{VOLUNTARY} + \beta_7 \text{INDEPENDENT} + \varepsilon \end{aligned} \quad (10.1)$$

$$\begin{aligned} \Delta \text{CONTRIBUTIONS}_a = & \beta_0 + \beta_1 \text{RATING}_a + \beta_2 \text{PARTNERSHIPS} \\ & + \beta_3 \text{RATING}_a \times \text{PARTNERSHIPS} + \beta_4 \text{TREND}_a \\ & + \beta_5 \text{CONTRIBUTIONS} + \beta_6 \text{VOLUNTARY}_a \\ & + \beta_7 \text{INDEPENDENT}_a + \varepsilon \end{aligned} \quad (10.2)$$

The argument has two key empirical implications, both of which concern the marginal effect of RATING_a on $\Delta \text{CONTRIBUTIONS}_a$. This is equal to $\beta_1 + \beta_3 \text{COMPETITION}$ in Equation (10.1) and $\beta_1 + \beta_3 \text{PARTNERSHIPS}$ in Equation (10.2), which entails that β_1 represents the marginal effect when $\text{COMPETITION} = 0$ in Equation (10.1) and $\text{PARTNERSHIPS} = 0$ in Equation (10.2). Since the argument posits that ratings do not influence resource flows when IOs are subject to little competition or possess weak partnerships, we should expect β_1 to be zero in both equations. The argument also implies that the marginal effect will become positive as COMPETITION and PARTNERSHIPS increase. Given that these variables are non-negative and RATING_a has a mean of zero (because it is standardized), IOs with below-average ratings have negative interaction term values that decrease with the variables, while IOs with above-average ratings have positive values that increase with them. Thus, the second implication is that β_3 will be positive in both equations.

Table 10.2 reports the results for Equation (10.1). In line with the argument, the coefficient on RATING_a is statistically significant at the 10 percent level in only one of the eight models, whereas the coefficient on $\text{RATING}_a \times \text{COMPETITION}$ is positive and significant in seven. The latter's lack of significance in Model (5) may be a consequence of the relatively late release of the Danish ratings (in 2012), which could result in attenuated estimates because resource flows in $FY_{i,t} - 1$ were already influenced by earlier assessments. As suggested in the introduction, however, they may also stem from characteristics of the ratings themselves or their creator (for instance, Denmark may have less credibility or influence in the donor community than

TABLE 10.2. Effect of performance ratings on changes in contributions as mediated by institutional competition (Equation 10.1)

Dependent Variable: $\Delta \text{CONTRIBUTIONS}_a (\log)$							
	(1)	(2)	(3)	(4)	(5)	(6)	(8)
	UK	AUS	NET	SWE	DEN	MOP(S)	MOP(R)
							AVG
RATING _{UK}	0.05 (0.05)						
RATING _{AUS}		0.14 (0.12)					
RATING _{NET}			-0.02 (0.05)				
RATING _{SWE}				0.13*** (0.04)			
RATING _{DEN}					0.06 (0.07)		
RATING _{MOP(S)}						-0.06 (0.06)	
RATING _{MOP(R)}							-0.09 (0.06)
RATING _{AVG}							-0.002 (0.06)
COMPETITION	-0.07 (0.05)	-0.11 (0.08)	0.01 (0.04)	-0.08** (0.03)	-0.04 (0.04)	-0.08 (0.05)	0.02 (0.05)
TREND _a	-0.67* (0.37)	-1.47* (0.80)	-1.18 (0.91)	-0.02 (0.45)	0.15 (1.34)	-0.35 (0.79)	0.1 (0.53)
							-0.41** (0.21)

Dependent Variable: Δ CONTRIBUTIONS _t (log)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	UK	AUS	NET	SWE	DEN	MOP(S)	MOP(R)	AVG
CONTRIBUTIONS (log)	-0.03 (0.05)	-0.08 (0.06)	-0.06 (0.04)	-0.06*** (0.02)	-0.02 (0.06)	-0.03 (0.05)	0.01 (0.04)	-0.05*** (0.02)
VOLUNTARY	0.003** (0.00)	0.0003 (0.00)	0.003* (0.00)	0.002 (0.00)	0.003 (0.00)	0.002 (0.00)	0.003*** (0.00)	0.002** (0.00)
INDEPENDENT	0.001 (0.00)	-0.01 (0.01)	0.001 (0.00)	-0.002* (0.00)	-0.001 (0.00)	0.002 (0.00)	0.0001 (0.00)	0.001 (0.00)
RATING _{UK} × COMPETITION	0.11*** (0.03)							
RATING _{AUS} × COMPETITION		0.07** (0.03)						
RATING _{NET} × COMPETITION			0.12*** (0.05)					
RATING _{SWE} × COMPETITION				0.05* (0.03)				
RATING _{DEN} × COMPETITION					-0.04 (0.06)			
RATING _{MOP(S)} × COMPETITION						0.13* (0.07)		
RATING _{MOP(R)} × COMPETITION							0.13** (0.06)	
RATING _{AVG} × COMPETITION								0.11*** (0.03)

Dependent Variable: $\Delta \text{CONTRIBUTIONS}_t(\log)$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	UK	AUS	NET	SWE	DEN	MOP(S)	MOP(R)	AVG
Constant	0.1 (0.37)	0.82** (0.40)	0.3 (0.22)	0.57*** (0.12)	0.03 (0.32)	0.08 (0.37)	-0.26 (0.27)	0.3 (0.18)
Observations	192	160	127	167	68	62	62	308
R ²	0.33	0.35	0.36	0.3	0.14	0.24	0.3	0.26
Adjusted R ²	0.3	0.32	0.32	0.26	0.04	0.14	0.21	0.24

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Notes: Ordinary least squares regressions with robust standard errors clustered by IO in parentheses.

the other assessor states). Encouragingly, the coefficient has the largest size and lowest p -value in the model with the aggregate index $RATING_{avg}$ (Model 8) – that is, the model with the highest sample size – providing evidence of the “triangulation” behavior described earlier.

The left panel of Figure 10.4 plots the estimated marginal effect of $RATING_{avg}$ on $\Delta CONTRIBUTIONS_a$ at different levels of COMPETITION. When IOs are subject to no competition, the effect is statistically indistinguishable from zero (as indicated by the intersection of $y = 0$ with the 90 percent confidence interval). As competition increases, however, it becomes positive and significant. More importantly, it becomes *substantively* significant: on average, a one-point increase in $RATING_{avg}$ (which ranges from -2.83 to 2.28) is associated with roughly a 25 percent rise in contributions when IOs have 5–10 competitors; a 35 percent rise when IOs have 10–20 competitors; and a 45 percent rise when IOs have more than 20 competitors.

The results for Equation (10.2), shown in Table 10.3, provide similarly strong support for the argument. The coefficient on $RATING_a$ fails to reach significance in all eight models, whereas the coefficient on $RATING_a \times PARTNERSHIPS$ is positive and significant in six. Similarly to before, the latter is non-significant only in Models (2) and (5) – in which ratings were first released in 2012 – and has the largest size and lowest p -value in Model (8). The right panel of Figure 10.4 displays the equivalent of the previous marginal effect plot for Model (8). As predicted, when partnerships are weak or non-existent ($PARTNERSHIPS < 1.5$ on a scale of 0 to 3.98), the estimated marginal effect of $RATING_{avg}$ on $\Delta CONTRIBUTIONS_a$ is indistinguishable from zero

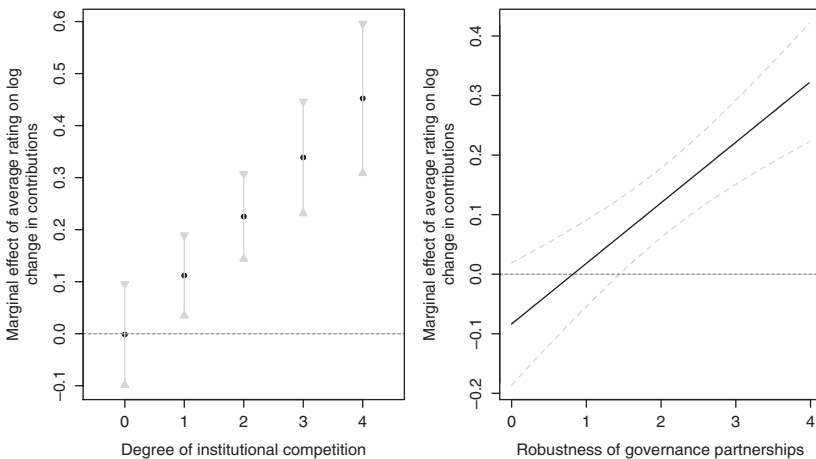


FIGURE 10.4. Marginal effects of average ratings on changes in contributions at different levels of institutional competition and partnership robustness.

Notes: The graphs are based on the results of Model (8) in Tables 10.2 (left graph) and 10.3 (right graph). The light gray lines represent 90 percent confidence intervals.

TABLE 10.3. Effect of performance ratings on changes in contributions as mediated by governance partnerships (Equation 10.2)

Dependent Variable: Δ CONTRIBUTIONS _a (log)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	UK	AUS	NET	SWE	DEN	MOP(S)	MOP(R)	AVG
RATING _{UK}	-0.04 (0.09)							
RATING _{AUS}		0.2 (0.17)						
RATING _{NET}			-0.23 (0.16)					
RATING _{SWE}				0.06 (0.07)				
RATING _{DEN}					0.04 (0.07)			
RATING _{MOP(S)}						-0.09 (0.06)		
RATING _{MOP(R)}							-0.15 (0.09)	
RATING _{AVG}								-0.08 (0.06)
PARTNERSHIPS	0.05 (0.04)	0.05 (0.06)	0.05 (0.04)	-0.01 (0.03)	0.01 (0.03)	0.02 (0.04)	0.04 (0.03)	0.03 (0.02)
TREND _a	-0.59 (0.36)	-1.42 (0.87)	-0.72 (0.74)	-0.2 (0.38)	0.27 (1.45)	-0.47 (0.76)	0.09 (0.61)	-0.41 (0.25)
CONTRIBUTIONS (log)	-0.04 (0.05)	-0.06 (0.06)	-0.04 (0.04)	-0.07*** (0.02)	-0.03 (0.08)	-0.07 (0.05)	-0.02 (0.03)	-0.04** (0.02)
VOLUNTARY	0.001 (0.00)	-0.002 (0.00)	0.002 (0.00)	0.002 (0.00)	0.003 (0.00)	0.002** (0.00)	0.004*** (0.00)	0.001 (0.00)
INDEPENDENT	0.0001 (0.00)	-0.01 (0.01)	0.002 (0.00)	-0.002* (0.00)	-0.0003 (0.00)	0.001 (0.00)	-0.001 (0.00)	0.0004 (0.00)

Dependent Variable: Δ CONTRIBUTIONS_a (log)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	UK	AUS	NET	SWE	DEN	MOP(S)	MOP(R)	AVG
RATING _{UK} × PARTNERSHIPS	0.08** (0.03)							
RATING _{AUS} × PARTNERSHIPS		-0.02 (0.08)						
RATING _{NET} × PARTNERSHIPS			0.12* (0.06)					
RATING _{SWE} × PARTNERSHIPS				0.04* (0.02)				
RATING _{DEN} × PARTNERSHIPS					-0.02 (0.02)			
RATING _{MOP(S)} × PARTNERSHIPS						0.05** (0.03)		
RATING _{MOP(R)} × PARTNERSHIPS							0.07* (0.04)	
RATING _{AVG} × PARTNERSHIPS								0.10*** (0.03)
Constant	0.17 (0.37)	0.69 (0.45)	0.09 (0.22)	0.61*** (0.16)	0.01 (0.34)	0.29 (0.33)	-0.15 (0.21)	0.32* (0.17)
Observations	192	160	127	167	68	62	62	308
R ²	0.3	0.32	0.36	0.27	0.1	0.19	0.23	0.22
Adjusted R ²	0.28	0.29	0.32	0.24	-0.002	0.09	0.13	0.21

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Notes: Ordinary least squares regressions with robust standard errors clustered by IO in parentheses.

(the lower 90 percent confidence bound intersects $y = 0$). When partnerships are robust, however, it is positive, significant, and substantively sizable. For IOs whose partnerships are in the upper quartile of the robustness distribution ($3.35 \leq \text{PARTNERSHIPS} \leq 3.98$), for instance, a one-point increase in $\text{RATING}_{\text{avg}}$ is associated with an average rise in contributions of roughly 25–30 percent. This is an impressive 35–40 percent higher than the corresponding figure when partnerships are at the bottom end of the distribution.

Tables A11–A12 in the online appendix display the results when the baseline equations are re-estimated with disaggregated standardized indicator scores from each assessment. Seventy-five of the 100 individual coefficients on the interaction terms are positive, of which one third are significant (only nine coefficients on RATING_a are positive and significant). There is little evidence that lower p -values are clustered in any of the six dimensions of performance listed in the second section, though (perhaps unsurprisingly) a higher proportion of the significant coefficients involve indicators of resource management and cost and value consciousness. In general, the fact that the results are stronger at the aggregate than at the indicator level suggests donors may be “triangulating” between different types of performance information *within* as well as *across* assessments.

Distinguishing Performance from Performance Indicators

One potential concern about the baseline models is that they fail to distinguish the effect of IO performance indicators from the effect of IO performance itself. It is conceivable, for instance, that the interaction-term coefficients are positive because ratings are capturing recent shifts in performance that have yet to influence resource flows or because performance has moved in the same direction as ratings since their release.

To address the first possibility, I employ a quasi-experimental design that involves expanding the sample to include unassessed IOs as a “control group.” I select such institutions by exploiting a sharp discontinuity in the rule determining the eligibility of IOs for inclusion in the United Kingdom’s assessment, namely, they must have received an average of at least one million British pounds in annual funding from the country’s Department for International Development (DFID) in the decade prior to the release of the first set of ratings in 2011.⁴⁶ Since IOs near each side of this threshold have similar characteristics, but varying levels of the treatment (see Tables A14–A16 for balance statistics), we can isolate the effect of performance ratings by restricting the sample to these institutions and estimating the following regression discontinuity (RD) models:

⁴⁶ Data on DFID funding come from the sources cited in fn. 33. The Australian government employs a similar threshold but does not disclose contributions to individual institutions. For a similar approach to distinguishing the effects of GPIs from the effects of underlying performance, see Roberts and Tellez this volume.

$$\begin{aligned}
\Delta \text{CONTRIBUTIONS}_{UK} = & \beta_0 + \beta_1 \text{POOLED}_{UK} + \beta_2 \text{COMPETITION} \\
& + \beta_3 \text{AUTHENTIC} + \beta_4 \text{POOLED}_{UK} \times \text{COMPETITION} + \beta_5 \text{POOLED}_{UK} \\
& \times \text{AUTHENTIC} + \beta_6 \text{COMPETITION} \times \text{AUTHENTIC} + \beta_7 \text{POOLED}_{UK} \\
& \times \text{COMPETITION} \times \text{AUTHENTIC} + \beta_8 \text{DFIDFUNDING} \\
& + \beta_9 \text{DFIDFUNDING}^2 + \beta_{10} \text{DFIDFUNDING}^3 + \varepsilon
\end{aligned} \tag{10.3}$$

$$\begin{aligned}
\Delta \text{CONTRIBUTIONS}_{UK} = & \beta_0 + \beta_1 \text{POOLED}_{UK} \\
& + \beta_2 \text{PARTNERSHIPS} + \beta_3 \text{AUTHENTIC} + \beta_4 \text{POOLED}_{UK} \\
& \times \text{PARTNERSHIPS} + \beta_5 \text{POOLED}_{UK} \times \text{AUTHENTIC} \\
& + \beta_6 \text{PARTNERSHIPS} \times \text{AUTHENTIC} + \beta_7 \text{POOLED}_{UK} \\
& \times \text{PARTNERSHIPS} \times \text{AUTHENTIC} + \beta_8 \text{DFIDFUNDING} \\
& + \beta_9 \text{DFIDFUNDING}^2 + \beta_{10} \text{DFIDFUNDING}^3 + \varepsilon
\end{aligned} \tag{10.4}$$

where POOLED_{UK} is a combined ratings measure equal to RATING_{UK} for the assessed IOs and a predicted rating based on the four control variables (which are strong predictors of RATING_{UK}) for the unassessed IOs;⁴⁷ AUTHENTIC is a dummy for whether an IO has been assessed; and DFIDFUNDING , the forcing variable, is an IO's average annual contributions from DFID over the period 2002–11 in millions of pounds (the polynomial terms address the possibility of a nonlinear relationship with the dependent variable). For the unassessed IOs, PARTNERSHIPS is measured in the same way as for the assessed IOs, while COMPETITION is measured by converting the number of IOs in the *Yearbook of International Organizations* database that share their “subject of activity” into the five-point scale described earlier.⁴⁸ If the authentic ratings have an independent conditional effect on resource flows – an effect distinct from that of IO performance itself – we should expect β_7 , the coefficient on the triple interaction term in each equation, to be positive; and β_4 , the marginal effect of the interaction between POOLED_{UK} and each conditioning variable on $\Delta \text{CONTRIBUTIONS}_{UK}$ when $\text{AUTHENTIC} = 0$, to be zero.

Table 10.4 presents the results for three different bandwidths: (1) IOs with a DFIDFUNDING value of less than 3.5 (the smallest bandwidth at which

⁴⁷ In a regression of RATING_{UK} on the controls, three of the four coefficients are significant at the 1 percent level (the other has a p -value of 0.24).

⁴⁸ *The Yearbook of International Organizations Online*. Available at <https://www.uia.org/yearbook>, accessed December 14, 2015. Both variables are measured as of 2010 (i.e. FY v_{10-1}). If an IO's subject of activity does not match one of the *Yearbook*'s near-thousand categories, COMPETITION takes a score of 0.

TABLE 10.4. Effect of pooled British ratings on changes in contributions as mediated by institutional competition and governance partnerships (Equations 10.3 and 10.4)

	Dependent Variable: $\Delta \text{CONTRIBUTIONS}_{UK} (\log)$											
	Cubic RD			Matching			Cubic RD			Matching		
	(1) DFIDFUND < 3.5	(2) DFIDFUND < 5	(3) Full Sample	(4) Full Sample	(5) DFIDFUND < 3.5	(6) DFIDFUND < 5	(7) Full Sample	(8) Full Sample				
POOLED _{UK}	0.31 (0.28)	0.31 (0.27)	0.31 (0.26)	0.34 (0.29)	0.71 (0.49)	0.69 (0.49)	0.57 (0.42)	0.32 (0.35)				
COMPETITION	-0.24 (0.19)	-0.24 (0.19)	-0.25 (0.19)	-0.27 (0.20)								
PARTNERSHIPS					-0.43 (0.29)	-0.41 (0.29)	-0.35 (0.25)	-0.14 (0.17)				
AUTHENTIC	-0.48 (0.52)	-0.31 (0.54)	-0.33 (0.41)	-0.36 (0.44)	-4.36** (1.68)	-0.74 (0.97)	-0.53 (0.60)	-0.12 (0.42)				
DFIDFUNDING	0.52 (0.45)	0.61 (0.39)	0.41 (0.01)	0.44 (0.01)	0.4 (0.41)	0.97** (0.44)	0.001 (0.01)					
DFIDFUNDING ²	-0.34 (0.34)	-0.44* (0.18)	0 (0.00)	0 (0.00)	0.42** (0.21)	-0.55*** (0.20)	0 (0.00)					
DFIDFUNDING ³	0.06 (0.08)	0.07*** (0.03)	0 (0.00)	0 (0.00)	-0.15*** (0.04)	0.08*** (0.03)	0 (0.00)					
POOLED _{UK} × COMPETITION	-0.17 (0.13)	-0.17 (0.13)	-0.17 (0.13)	-0.21 (0.14)								
POOLED _{UK} × PARTNERSHIPS					-0.35 (0.22)	-0.34 (0.22)	-0.29 (0.19)	-0.25 (0.16)				

Dependent Variable: $\Delta \text{CONTRIBUTIONS}_{UK} (\log)$

	Cubic RD			Matching			Cubic RD			Matching		
	(1) DFIDFUND < 3.5	(2) DFIDFUND < 5	(3) Full Sample	(4) Full Sample	(5) DFIDFUND < 3.5	(6) DFIDFUND < 5	(7) Full Sample	(8) Full Sample				
POOLED _{UK} × AUTHENTIC	-0.28 (0.28)	-0.34 (0.29)	-0.23 (0.27)	-0.27 (0.30)	-2.39*** (0.91)	-0.6 (0.53)	-0.56 (0.43)	-0.33 (0.35)				
COMPETITION × AUTHENTIC	0.16 (0.20)	0.21 (0.20)	0.23 (0.20)	0.27 (0.21)								
POOLED _{UK} × COMPETITION × AUTHENTIC	0.62*** (0.17)	0.29* (0.15)	0.23* (0.13)	0.24* (0.14)								
PARTNERSHIPS × AUTHENTIC					1.24** (0.49)	0.34 (0.30)	0.35 (0.26)	0.13 (0.17)				
POOLED _{UK} × PARTNERSHIPS × AUTHENTIC					0.82** (0.32)	0.33 (0.23)	0.37* (0.19)	0.22* (0.13)				
Constant	0.39 (0.41)	0.38 (0.41)	0.48 (0.40)	0.48 (0.43)	0.77 (0.62)	0.69 (0.62)	0.69 (0.58)	0.25 (0.42)				
Observations	200	225	357	517	200	225	357	517				
R ²	0.09	0.1	0.12	0.09	0.13	0.13	0.14	0.09				
Adjusted R ²	0.04	0.06	0.09	0.08	0.08	0.09	0.12	0.08				

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Notes: Robust standard errors clustered by IO in parentheses. In Models (4) and (8), assessed (AUTHENTIC = 1) and unassessed (AUTHENTIC = 0) observations are genetically matched on the conditioning and control variables, with missing values (which are not permitted) filled in using multiple imputation.

there are enough assessed observations to estimate the equations); (2) IOs with a DFIDFUNDING value of less than 5; and (3) the full sample.⁴⁹ In addition, it reports the results of an alternative strategy for improving balance across the treated and control groups, namely, matching assessed and unassessed observations on the conditioning and control variables.⁵⁰ Consistent with expectations, seven of the eight coefficients on the triple interaction terms are positive and significant, with the lowest *p*-values in the models with the smallest bandwidth (Models 1 and 5), in which the assessed and unassessed IOs are most similar. By contrast, every coefficient on POOLED_{UK} × COMPETITION and POOLED_{UK} × PARTNERSHIPS is nonsignificant and negative.⁵¹ Comparing the estimated marginal effect of POOLED_{UK} on ΔCONTRIBUTIONS_{UK} at different levels of the conditioning variables for assessed and unassessed IOs in the DFIDFUNDING < 3.5 bandwidth indicates that the treatment makes a major substantive difference: contributions rise by roughly 45 percent for assessed IOs versus -15 percent for unassessed IOs with each one-point increase in COMPETITION; and 45 percent versus -35 percent with each one-point increase in PARTNERSHIPS.

The second possibility mentioned above – that performance has moved in the same direction as ratings since their release – is difficult to address directly. However, it can be substantially mitigated by making a small adjustment to the dependent variable, namely, restricting the post-assessment period to FY v_{IO} (i.e. changing the variable's numerator to CONTRIBUTIONS _{v_{IO}}). This is because performance exhibits relatively little change from year to year, and it is unlikely that a high proportion of IOs would coincidentally experience shifts in the same direction as ratings during the year of the latter's release. The results of the modified baseline equations, shown in Tables A21 and A22, continue to support the argument, though are marginally weaker than those in Tables 10.2 and 10.3 – as should be expected, given the smaller sample size and the fact that many donors are likely to be either unaware of or unable to respond to ratings during FY v_{IO} .⁵²

Additional Robustness Checks

The findings remained consistent with the argument throughout a number of alternative specifications of the baseline models, the results of which are also

⁴⁹ Table A13 provides a list of the unassessed IOs – which exclude those that have been rated in any of the other assessments – and their values of DFIDFUNDING. The third column of the list of assessed IOs in the chapter appendix shows their DFIDFUNDING values, which range from 1.19 to 78,018.13 and have a mean of 1,880.16.

⁵⁰ I employ genetic matching, filling in missing values – which are not permitted – using multiple imputation. Tables A17 and A18 display post-matching balance statistics.

⁵¹ As shown in Tables A19 and A20, these results are robust to the use of lower order polynomials.

⁵² Importantly, the results are stronger in the models with larger sample sizes (in particular Model 8).

reported in the online appendix.⁵³ First, I included FY $v_{IO} + i$ fixed effects to address temporal heterogeneity, a serious possibility given that the dataset covers the financially turbulent period following the global economic crisis (Tables A23 and A24). Second, I controlled for four additional variables: AGE, an IO's log age in FY $v_{IO} - 1$; SCOPE, the number of issue areas in which an IO operates in FY $v_{IO} - 1$; REGIONAL, a dummy for whether an IO has a regional focus in FY $v_{IO} - 1$; and DIVERGENCE, the variance of an IO's standardized mean indicator scores in each assessment (since resource flows may be more responsive to ratings when there is greater agreement among assessors) (Tables A25 and A26). Third, I included both sets of conditioning variables and interaction terms in the same equation (Table A27). Fourth, I excluded influential observations from the analyses, which were identified using the widely used Cook's Distance metric (Tables A28 and A29).⁵⁴

A final set of checks involved experimenting with alternative measures of key variables in the analysis. To address the possibility that resource flows in FY $v_{IO} - 1$ are anomalous relative to general pre-assessment funding levels, I constructed modified measures of the dependent variable in which the denominator is changed to $\sum_{j=1}^3 \text{CONTRIBUTIONS}_{v_{IO},j}/3$ (Tables A30 and A31) and $\sum_{j=1}^5 \text{CONTRIBUTIONS}_{v_{IO},j}/3$ (Tables A32 and A33), i.e. the pre-assessment period is expanded to three and five FYs, respectively. In addition, I employed three different measures of institutional competition, the first two based on the issue-specific determinants of the variable discussed in the third section and the third based on the measure developed for unassessed IOs in the previous subsection: (1) a proxy for expertise-based barriers to entry that measures whether an IO performs specialized technical assistance functions for its members in FY $v_{IO} - 1$ (EXPERTISE) (Table A34); (2) a proxy for network effects-based economies of scale that measures whether an IO engages in standard-setting activities in FY $v_{IO} - 1$ (STANDARDS) (Table A35); and (3) the log number of IOs in the *Yearbook of International Organizations* database that share an IO's subject of activity in FY $v_{IO} - 1$ (SHARED) (Table A36).⁵⁵

CONCLUSION

As a form of social knowledge produced by respected and authoritative actors, IO performance indicators highlight and shape shared understandings of the

⁵³ The interaction-term coefficients marginally exceeded the $p < 0.1$ significance threshold in some models with lower sample sizes and observation-to-variable ratios.

⁵⁴ Observations were excluded if their Cook's Distance exceeded $4/n$ (where n is the analysis sample size).

⁵⁵ The interaction-term coefficients in Tables A33 and A34 are negative because barriers to entry and economies of scale are inversely related to competition. The results of these models provide stronger support for the unconditional GPA influence hypothesis, but should be treated with some caution because EXPERTISE and STANDARDS are less direct measures of competition.

behavior and impact of such institutions, enhancing the salience of performance as a basis for making multilateral funding decisions. I have argued, however, that performance ratings do not influence resource flows to IOs under all circumstances; rather, their impact is conditional on the nature of the relationship between IOs and other actors within their policy space. Specifically, they only exert influence when IOs (1) are subject to a high degree of institutional competition and (2) possess robust governance partnerships with non-state actors. A combination of qualitative and quantitative evidence from a variety of original sources has provided consistent empirical support for the argument.

In addition to furthering our understanding of the sources – and limits – of assessment power in world politics, the findings have implications for other kinds of GPIs with the potential to influence material resource flows to target entities, such as indicators of democracy, governance, corruption, and business conditions.⁵⁶ They suggest, for instance, that such GPIs will have a greater impact on resource flows when there is a sizable pool of target entities that possess similar characteristics and could thus serve as substitutes for one another. Thus, we might expect assessments of a state's business conditions or governance to be less likely to affect foreign direct investment (FDI) inflows if it possesses a particularly rare natural resource or large internal market. The argument also implies that GPIs will have stronger material effects when assessed entities have close operational ties with actors capable of influencing resource holders – or with resource holders themselves. Economies that are highly integrated with the home countries of foreign investors, for example, may be more likely to receive a boost in FDI inflows when their political or business climate receives a positive assessment.

The findings also have implications for understanding the *non-material* effects of GPIs. In addition to funding levels, IO performance indicators are likely to influence outcomes such as the size of an IO's membership, the degree of (state and non-state) compliance with its policies, and the extent of its soft power and informal influence in the international system. If my argument is correct, we should expect these variables to be more responsive to performance ratings when IOs face few competitors and have powerful allies above and below the state. By a similar logic, we might expect assessments of a state's level of democracy or internal security to have a weaker impact on outcomes such as its ability to form alliances and engage in institutionalized cooperation when it occupies a strategically important position in the international system (for instance, due to its geographical location or capabilities) and has close political, economic, and social ties with potential cooperators. These and the above possibilities point to relational analyses of the material and non-material effects of GPIs as a fruitful area for further research.

⁵⁶ See Kelley, Simmons, and Doshi; Lee and Matanock; Morse; and Roberts and Tellez, this volume.

Another promising research avenue concerns the sources of variation in governance partnerships. While my argument sheds light on the factors affecting institutional competition, it does not directly address the question of why some IOs are able to form more robust partnerships than others. A comprehensive answer to this question is beyond the scope of this chapter, though a few possibilities are worth mentioning. First, some policy spaces are simply populated by fewer non-state actors than others, for instance, because the issue in question has little popular resonance or is associated with severe collective action problems.⁵⁷ Second, partnership formation may be influenced by the openness of an IO's policymaking process to external actors (in part a function of institutional design), which affects its opportunities to identify and interact with potential partners.⁵⁸ Finally, non-state actors sometimes play an important role in the establishment of IOs, forging informal ties with such institutions that may provide the basis for future operational collaboration.⁵⁹ In other words, variation in partnership robustness may reflect historical legacies of institutional creation. Developing and testing a full theory of partnership formation is an important task for future research.

References

- Aaron, Joshua R., Amy McMillan, and Brandon N. Cline. 2012. Investor Reaction to Firm Environmental Management Reputation. *Corporate Reputation Review* 15 (4):304–18.
- Abbott, Kenneth W., Philipp Genschel, Duncan Snidal, and Bernhard Zangl, eds. 2015. *International Organizations as Orchestrators*. Cambridge: Cambridge University Press.
- Abbott, Kenneth W., Jessica F. Green, and Robert O. Keohane. 2016. Organizational Ecology and Institutional Change in Global Governance. *International Organization* 70 (2):247–77.
- Abbott, Kenneth W., and Duncan Snidal. 1998. Why States Act through Formal International Organizations. *Journal of Conflict Resolution* 42 (1):3–32.
- Ackrill, Robert, and Adrian Kay. 2006. Historical-Institutionalist Perspectives on the Development of the EU Budget System. *Journal of European Public Policy* 13 (1):113–33.
- Alter, Karen J., and Sophie Meunier. 2009. The Politics of International Regime Complexity. *Perspectives on Politics* 7 (1):13–24.

⁵⁷ Abbott et al. 2015 make a similar conjecture. In addition, they posit that partnerships are more likely to be formed when IOs lack the capabilities to achieve their goals, occupy a focal institutional position in their issue area, have entrepreneurial cultures, and are subject to weak oversight by states.

⁵⁸ See Tallberg et al. 2013.

⁵⁹ Johnson 2014.

- Bisbee, James H., James R. Hollyer, B. Peter Rosendorff, and James Raymond Vreeland. this volume. The Millennium Development Goals and Education: Accountability and Substitution in Global Indicators. In *The Power of Global Performance Indicators*, edited by Judith Kelley and Beth A. Simmons. New York: Cambridge University Press.
- Brammer, Stephen, Chris Brooks, and Stephen Pavelin. 2009. The Stock Performance of America's 100 Best Corporate Citizens. *The Quarterly Review of Economics and Finance* 49 (3):1065–80.
- Broz, J. Lawrence, and Michael B. Hawes. 2006. Congressional Politics of Financing the International Monetary Fund. *International Organization* 60 (2):367–99.
- Busby, Joshua William. 2007. Bono Made Jesse Helms Cry: Jubilee 2000, Debt Relief, and Moral Action in International Politics. *International Studies Quarterly* 51 (2):247–75.
- Cameron, Kim S. 1978. Measuring Organizational Effectiveness in Institutions of Higher Education. *Administrative Science Quarterly* 23 (4):604–32.
- Cooley, Alexander, and Jack Snyder, eds. 2015. *Ranking the World: Grading States as a Tool of Global Governance*. Cambridge: Cambridge University Press.
- Davis, Kevin E., Angelina Fisher, Benedict Kingsbury, and Sally Engle Merry. 2012. *Governance by Indicators: Global Power through Classification and Rankings*. Oxford: Oxford University Press.
- Department for International Development (United Kingdom). 2013. *Multilateral Aid Review Update*. London: DFID.
- Department of Foreign Affairs and Trade (Australia). 2013. Australia's International Development Assistance Program 2013–14. Statement by Senator the Hon. Bob Carr, Minister for Foreign Affairs, May 14, 2013, Canberra.
- Etzioni, A. 1964. *Modern Organizations*. Englewood Cliffs, NJ: Prentice-Hill.
- Food and Agriculture Organization. 2007. *Report of the Independent External Evaluation of the Food and Agriculture Organization of the United Nations (FAO)*. Conference document C 2007/7A.1-Rev.1, 34th Session. Rome.
- Frey, Bruno S. 2008. Outside and Inside Competition for International Organizations – From Analysis to Innovations. *The Review of International Organizations* 3 (4):335–50.
- Gavas, Mikaela. 2012. Reviewing the Evidence: How Well Does the European Development Fund Perform? ODI Discussion Paper. London: Overseas Development Institute.
- Hawkins, Darren G., David A. Lake, Daniel L. Nielson, and Michael J. Tierney, eds. 2006. *Delegation and Agency in International Organizations*. Cambridge: Cambridge University Press.
- Honig, Dan, and Catherine Weaver. this volume. A Race to the Top?: The Aid Transparency Index and the Normative Power of Global Performance Indicators. *Annual Meeting of the American Political Science Association*.
- Hooghe, Liesbet, Gary Marks, Tobias Lenz, Jeanine Bezuijen, Besir Ceka, and Svet Derderyan. 2017. *Measuring International Authority: A Postfunctionalist Theory of Governance, Volume III*. Oxford: Oxford University Press.
- Hoole, Francis W., David H. Handley, and Charles W. Ostrom. 1979. Policy-Making Models, Budgets and International Organizations. *The Journal of Politics* 41 (3):923–32.

- Hoole, Francis W., Brian L. Job, and Harvey J. Tucker. 1976. Incremental Budgeting and International Organizations. *American Journal of Political Science* 20 (2):273–301.
- Johnson, Tana. 2014. *Organizational Progeny: Why Governments Are Losing Control over the Proliferating Structures of Global Governance*. Oxford: Oxford University Press.
- Jupille, Joseph, Walter Mattli, and Duncan Snidal. 2013. *Institutional Choice and Global Commerce*. Cambridge: Cambridge University Press.
- Kelley, Judith G., and Beth A. Simmons. this volume. Introduction: The Power of Global Performance Indicators. Chapter 1 in *The Power of Global Performance Indicators*.
- Kelley, Judith G., Beth A. Simmons, and Rush Doshi. this volume. The Power of Ranking: The Ease of Doing Business and Global Regulatory Behavior. Chapter 2 in *The Power of Global Performance Indicators*.
- Lavelle, Kathryn C. 2011. Multilateral Cooperation and Congress: The Legislative Process of Securing Funding for the World Bank. *International Studies Quarterly* 55 (1):199–222.
- Lee, Melissa M., and Aila M. Matanock. this volume. Third Party Policymakers and the Limits of the Influence of Indicators. Chapter 11 in *The Power of Global Performance Indicators*. Cambridge: Cambridge University Press.
- Lipsy, Phillip Y. 2015. Explaining Institutional Change: Policy Areas, Outside Options, and the Bretton Woods Institutions. *American Journal of Political Science* 59 (2):341–56.
- Mattli, Walter, and Jack Seddon. 2015. Orchestration along the Pareto frontier: winners and losers. In *International Organizations as Orchestrators*, edited by Kenneth W. Abbott, Philipp Genschel, Duncan Snidal, and Bernhard Zangl. Cambridge: Cambridge University Press.
- Merry, Sally Engle, Kevin E. Davis, and Benedict Kingsbury, eds. 2015. *The Quiet Power of Indicators: Measuring Governance, Corruption, and Rule of Law*. Cambridge: Cambridge University Press.
- Ministerie van Buitenlandse Zaken (Netherlands). 2013. Kamerbrief over scorecards internationale organisaties. Den Haag.
- Ministry for Foreign Affairs (Denmark). 2013. Danish Multilateral Development Cooperation Analysis: An Assessment of Denmark's Multilateral Engagement in Light of the Right to a Better Life, the Strategy for Danish Development Cooperation. Copenhagen.
- Ministry for Foreign Affairs (Sweden). 2011. Swedish Assessment of Multilateral Organisations: Summary 2011. Stockholm.
- Morse, Julia C. this volume. Blacklists, Market Enforcement, and the Global Regime to Combat Terrorist Financing. *Annual Meeting of the American Political Science Association*.
- Multilateral Organization Performance Assessment Network. 2015. Annual Report 2014. Paris.
- Murguía, Juan M., and Sergio H. Lence. 2015. Investors' Reaction to Environmental Performance: A Global Perspective of the Newsweek's "Green Rankings." *Environmental and Resource Economics* 60 (4):583–605.
- Nielson, Daniel, and Michael J. Tierney. 2003. Delegation to International Organizations: Agency Theory and World Bank Environmental Reform. *International Organization* 57 (2):241–76.

- Pevehouse, Jon, Timothy Nordstrom, and Kevin Warnke. 2004. The Correlates of War 2 International Governmental Organizations Data Version 2.0. *Conflict Management and Peace Science* 21 (2):101–19.
- Pollack, Mark A. 1997. Delegation, Agency, and Agenda Setting in the European Community. *International Organization* 51 (1):99–134.
- Price, James L. 1972. The Study of Organizational Effectiveness. *The Sociological Quarterly* 13 (1):3–15.
- Roberts, Jordan, and Roberts Fernando Tellez. this volume. Freedom House's Scarlet Letter: Assessment Power through Transnational Pressure.
- Singer, J. David. 1961. *Financing International Organization: The United Nations Budget Process*. The Hague: Nijhoff.
- Tallberg, Jonas, Thomas Sommerer, Theresa Squatrito, and Christer Jönsson. 2013. *The Opening Up of International Organizations: Transnational Access in Global Governance*. Cambridge: Cambridge University Press.