



# Explaining the political use of evaluation in international organizations

Steffen Eckhard<sup>1</sup> · Vytautas Jankauskas<sup>1,2</sup>

Published online: 28 August 2020  
© The Author(s) 2020

## Abstract

Despite a growing literature on the politics of evaluation in international organizations (IOs) and beyond, little is known about whether political or administrative stakeholders indeed realize ex-ante political interests through evaluations. This is, however, especially important considering the booming business of evaluation and the proliferation of institutional assessments both in domestic and international politics. We argue that formally independent IO evaluation units informally orientate towards either member states or the IO administration, depending on who controls the unit's budget, staff, and agenda resources. This should enable either actor to also use evaluation results along pre-defined strategic interests. Interview data gathered among evaluators, secretariat officials, and member state representatives of six IOs support the expected pattern, highlighting striking differences in the orientation of evaluation staff and evaluation use. Findings challenge the technocratic, apolitical image of evaluation, offering practical and theoretical implications for future research.

**Keywords** International organization · International public administration · Evaluation · United Nations · Evidence-based policy-making

## Introduction

In public policy-making, evaluation is typically perceived as a functional tool in the final phase of a cyclical process (Anderson 1975; Brewer 1974; Howlett et al. 2009; Lasswell 1956). Evaluation is thought to inform policy-makers about past and present experiences, allowing for policy adjustment, learning, and accountability. This essentially follows a functionalist logic, where policy-makers seek effectiveness and efficiency, evaluation being one of the measures they take. Yet, in the political reality of public organizations,

---

✉ Steffen Eckhard  
steffen.eckhard@uni-konstanz.de

Vytautas Jankauskas  
v.jankauskas@gsi.uni-muenchen.de

<sup>1</sup> University of Konstanz, Universitaetsstr. 10, 78467 Constance, Germany

<sup>2</sup> Geschwister-Scholl-Institut für Politikwissenschaft, Ludwig-Maximilians-Universität München, Oettingenstraße 67, 80538 Munich, Germany

evaluation finds itself caught between stakeholder struggles for interests and power (Azzam 2010; Bjornholt and Larsen 2014; Morris and Clark 2013; Pleger et al. 2017; Taylor and Balloch 2005; Wildavsky 1972; Weiss 1998). For instance, evaluation results may be helpful in justifying actors' bargaining positions on policy decisions: "Whenever an evaluation affects the future allocation of resources and, hence, a change in power relationships, it is a political activity" (Wergin 1976, p. 76).

Many authors lament these "politics of evaluation" (e.g. Banner 1974; Taylor and Balloch 2005; Weiss and Jordan 1976), suggesting that evaluation results can become "ammunition in political battles" (Schoenefeld and Jordan 2019, p. 377) or that stakeholders put pressure on evaluators "to misrepresent findings" (Pleger et al. 2017, p. 316). Van Voorst and Mastebroek (2019) show that the quality of policy evaluation systematically decreases when additional political stakeholders get involved. Despite such insights, there is very little comparative evidence detailing whether policy-makers indeed realize ex-ante political interests through evaluation. Taking it from there, this paper *asks whether evaluation systematically serves ex-ante political interests of policy actors that go beyond the traditional duet of learning and accountability*. Such a political use of evaluation is defined as situations in which actors refer to evaluations (their findings, processes, or recommendations) to realize own political interests in competition with others.

The empirical focus is on international organizations (IOs) in the United Nations (UN) system, which offers an important analytical advantage. Comparative studies of evaluation use in domestic settings are challenging because of idiosyncrasies of individual agencies or political systems. IOs, by contrast, offer a class of comparable cases. UN system IOs all abide by the same evaluation norms and guidelines of the UN Evaluation Group (UNEG) which rules out a range of confounding factors.<sup>1</sup>

To identify political interests that could be realized using evaluation, we apply the classical lenses of principal-agent theory. On the one hand, member states are the principals in IOs, who take all policy decisions (Rittberger et al. 2019). Their political interests relate to internal dynamics among the state collective and their relationship with the IO administration. Evaluation can help states in *intergovernmental bargaining* and in exercising *agency control*. On the other hand, a growing literature emphasizes that IO secretariats (International Public Administrations—IPAs) act as independent actors in international policy processes (Bauer et al. 2017; Biermann and Siebenhüner 2009; Eckhard and Ege 2016; Patz and Goetz 2019; Knill et al. 2019; Johnson and Urpelainen 2014).<sup>2</sup> Evaluation can be politically useful for IPAs, to *justify past and future action*, or to *steer the organization internally*. For instance, studying the IMF, Hinterleitner, Sager and Thomann (2016, p. 564) found that "evaluations are a relatively obvious way for the IMF to effectively exert indirect influence on member states via its surveillance activities".

Existing research highlights that powerful stakeholders can use the institutional context surrounding evaluations to exercise their influence (Azzam 2010; Højlund 2014; Raimondo 2018). In IOs, this could be done through the control over IO evaluation systems resources,

<sup>1</sup> UNEG defines evaluation as "an assessment, conducted as systematically and impartially as possible, of an activity, project, programme, strategy, policy, topic, theme, sector, operational area or institutional performance" (UNEG 2016).

<sup>2</sup> IPAs are "bodies with a certain degree of autonomy, staffed by professional and appointed civil servants who are responsible for specific tasks and who work together following the rules and norms of the IO in question" (Bauer et al. 2017, p. 2). For member states, it is primarily their delegated representatives who constitute IO governing bodies.

i.e. evaluation units' staff, budget, and agenda (see Eckhard and Jankauskas 2019). The main expectation therefore is *that the political use of evaluation should differ systematically between IOs whose evaluation resources are either controlled by the IPA or by IO member states*. Such political use should become visible at two stages: We hypothesize that (1) independent IO evaluation units—as key actors managing centralized evaluations in IOs—should orient to the dominant stakeholder; and (2) the use of evaluation in IO policy-making should systematically reflect the typical political interests of the dominant stakeholder.

Empirically, we realize a small-N comparison. The sample includes nine IOs (IAEA, IOM, ILO, UNICEF, UNDP, UNEP, UNHCR, UNESCO, and WHO), six of which are analysed.<sup>3</sup> They fall into three groups, distinguished according to who controls evaluation system resources (the IPA, member states, or both). We draw on 35 qualitative expert interviews with senior officials in the IO administration, evaluation units, and member state representations. Data demonstrate that differences in the *control over evaluation resources* (iV) link to differences in the *orientation of IO evaluation units* (dV<sub>1</sub>) to either IPAs or member states and also to differences in the *alignment of evaluation use* (dV<sub>2</sub>) towards the political interests of either IPAs or member states.

The study is the first to present comparative empirical data on political evaluation use in IOs. It shows that the same evaluation tool, governed by the same UN system-wide evaluation standards, plays a very different role in IO policy-making, depending on who controls evaluation units. While these findings do not claim that evaluation reports as such are biased, we demonstrate that their political use differs systematically. These findings concur with others who argue that the idea of evidence-based policy-making should not be taken for granted (Cooley and Snyder 2015; Hinterleitner et al. 2016; Højlund 2015; Merry 2011; Porter 1995; van Voorst and Mastenbroek 2019). We do not suggest to refrain from using evaluation in public management, but the revealed politics surrounding seemingly functional tools like evaluation should be acknowledged and factored into our understanding of policy-making processes (for a related discussion, see Fforde 2019; Perl et al. 2018). For now, such a political understanding of evaluation appears uncommon among IO practitioners.

In the following sections, we first review the literature on IOs to derive expectations about evaluation stakeholder interests and the associated motives for political evaluation use. We then discuss how these dynamics can be observed empirically, introduce our research design and methods, present the empirical data, and discuss our findings as well as theoretical and practical implications.

---

<sup>3</sup> We deliberately exclude three of the IOs from further analysis to ensure interviewees' anonymity (see section on research design for a detailed explanation). The acronyms refer to the International Atomic Energy Agency (IAEA), International Organization for Migration (IOM), International Labour Organization (ILO), United Nations Children's Fund (UNICEF), United Nations Development Programme (UNDP), United Nations Environment Programme (UNEP), United Nations High Commissioner for Refugees (UNHCR), United Nations Educational, Scientific and Cultural Organization (UNESCO), and World Health Organization (WHO).

**Table 1** Overview of theorized principal-agent dynamics in IOs

P-A dynamic	Principal (MS)	Agent (IPA)
Within actor	Collective principal dynamics	Collective agent dynamics
Between actors	P-A control dynamics (bureaucratic drift)	Bureaucratic influence and justification dynamics

## Theorizing political interests and the political use of evaluation in IOs

Evaluation researchers have conceptualized the use of evaluation in three ways: “[e]valuations could be used (a) instrumentally, to give direction to policy and practice; (b) politically or symbolically, to justify pre-existing preferences and actions, and (c) conceptually, to provide new generalizations, ideas, or concepts” (Weiss et al. 2005, p. 13). Whereas (a) and (c) follow a functional rationale, the political (or symbolic) evaluation use forms the focus of this article.

In this regard, previous studies indicate that evaluations in IOs are politically contested (Højlund 2015, p. 35; Raimondo 2018, p. 32; van Voorst and Mastenbroek 2019; Weaver 2010, p. 366). If evaluation is indeed used politically, i.e. “in the battle for influence and interest promotion” (Bjornholt and Larsen 2014, p. 407), the first step of the analysis is to systematically unpack the pre-existing political interests of key evaluation stakeholders in IOs and to assess how evaluation may serve them.

Due to their direct embeddedness in decision-making processes, the main evaluation stakeholders in IOs are member states on the one hand and IO administrations on the other.<sup>4</sup> We apply the classical lenses of principal-agent (P-A) theory to distinguish between member states as the principal and IO bureaucracies (IPAs) as the agent. This way and by drawing on a wide range of literature, we are able to outline general P-A-dynamics referring to dynamics *within* the two actors (among member states and within the IPA) as well as *between* the actors (member states interests vis-à-vis the IPA and vice versa). These P-A-dynamics, in turn, generate political interests which exist independently from issue area or IO characteristics. The following sections describe the different interest dynamics from each of the quadrants (Table 1) and how evaluation can help actors utilizing them.

First, IO member states operate as a *collective principal* with heterogeneous political interests (Hawkins et al. 2006). Although states formally resolve their disagreements during voting in the policy-making forum of an IO, they are still constantly fighting over IO policies (Koremenos et al. 2001). In this regard, evaluation might be a strategic tool in states’ negotiations. Its results can be used to strengthen their own argumentation or shift blame to others (Chelimsky 1987). Given that evidence is power (see Botterill and Hindmoor 2012), convincing other state counterparts should become easier when referring to arguably objective findings rather than ideological standpoints. Furthermore, individual member states were observed for seeking influence by manipulating the bureaucratic agent—the IPA (Urpelainen 2012; Dijkstra 2015). For instance, powerful member states would push their unilateral agenda by rewarding or punishing influential staff within

<sup>4</sup> This is analogous to key evaluation stakeholders at the domestic level: political decision-makers and bureaucratic servants (Bjornholt and Larsen 2014, p. 408; Leeuw and Furubo 2008; The LSE GV314 Group 2014).

the secretariat to serve their interests in policy-making or implementation (Voeten 2008; Streck 2001). Against this backdrop, other states could use evaluation to investigate or prevent such practices. Thus, evaluation could be used as a safeguard to counter unilateral influence within the member states collective.

Second, IPAs have also been recognized as *collective agents*. Graham (2014) argued that IO bureaucracies should be treated as plural actors with structurally fragmented organizational units. Hanrieder (2015) also outlined the concept of fragmentation in the context of bureaucratic complexity, arguing that IO subunits, such as regional or country offices, often have their own interests and power. IO complexity means that “[t]he greater the extent of these subunit authorities, the more fragmented and the less hierarchical is an IO” (Hanrieder 2015, p. 34). It might thus become a fundamental challenge for an IO’s management to keep the competing organizational parts together. Evaluation can be used in this sense as a tool to vertically steer the organization. It provides information about the performance of organizational subunits, allowing senior management to overcome informational asymmetry within the fragmented structure. Just as member states may use evaluation to avoid agency slack from the bureaucracy (see below), the administration can also use evaluation to strengthen itself by preventing drift within its own ranks.

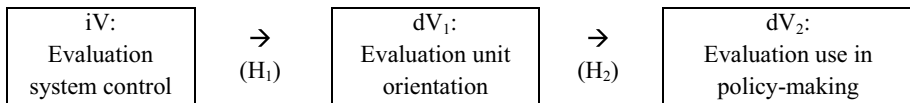
Third, turning to the behavioural dynamics *between* the actors, the extensive principal-agent literature demonstrates that the asymmetric structure of delegation relationships allows the possibility of agency losses, so that *member states have to employ control over IPAs* (Hawkins et al. 2006; Kassim and Menon 2003; Nielson and Tierney 2003). While the principal seeks to control its agent to avoid unwanted behaviour, the latter may have their own preferences and seek to escape the former’s oversight. da Conceição-Heldt (2013, p. 24) described it as “agents’ ability to act independently of their principals and to overreach their delegated authority”. Evaluation may hence serve as an ex-post control instrument for member states to contain bureaucratic influence. Ideally, evaluation reports help to reduce the informational gap by red-flagging unwanted IPA behaviour.

Fourth, evaluation may also serve the exact opposite purpose—for the *IPA to exercise bureaucratic influence* if it wants to do less or differently than asked by the member states (Elsig 2011; Vaubel et al. 2007). A number of studies show that IPAs are highly innovative and entrepreneurial in employing various administrative tools and tactics to increase own autonomy and influence policy-making (Patz and Goetz 2019; Knill et al. 2019; Biermann and Siebenhüner 2009; Bauer et al. 2017; Johnson and Urpelainen 2014). Consequently, evaluation might be a tool for IPAs to justify such initiatives. For instance, IPAs may refer to evaluations *vis-à-vis* the member states for resource mobilization strategies, favourable agenda-setting, mandate expansion, or to justify past actions (Easterly 2006; O’Brien et al. 2010; Weiss 1998). This relates to Goffman’s “impression management” (1959), where actors “present different aspects of themselves to suit the particular audiences they are faced with”. For instance, Hayward and colleagues (The LSE GV314 Group 2014, p. 224) demonstrated that UK civil servants put significant efforts to shape evaluation results “making them ‘look good’ or minimizing criticism of their policies”.

Considering the above, evaluation in IOs can be used in many more ways than only for the functional purpose of accountability and learning. Of course, the functional and political imperatives are not necessarily inversely related. We might observe both functional and strategic/political evaluation use in the same IO, even in the same evaluation. However, the aim of this paper is to identify the emergence and existence of the latter. Based on the theoretical identification of different motives for political evaluation use (see Table 2), the following section turns to the explanation of under what conditions which kind of political use should prevail.

**Table 2** Conceptualization of political evaluation use for each actor based on principal-agent dynamics

Principal-agent dynamic	Actor political interest	Political evaluation use
Member states		
Collective principal	Justifying own interests and preventing unilateral influence	Evaluation used in negotiations among states and to counter unilateral influence
P-A control	Controlling the agent	Evaluation used to contain IPA influence/agency drift
International public administration		
Collective agent	Containing organizational fragmentation	Evaluation used to vertically steer IPA and contain fragmentation
Bureaucratic influence	Justifying past or future action	Evaluation used to justify IPA initiatives and past action



**Fig. 1** The theorized causal mechanism

## Explaining political evaluation use in IOs

For the purpose of this research, we focus on IOs' centralized evaluation function, which covers evaluations produced (or managed) by IO evaluation units. This scope condition excludes decentralized evaluations which are mostly routinized project-level studies, usually conducted by responsible IPA departments. Centralized evaluations are mostly high-level ex-post assessments that “generally support overall corporate-level policy and strategic decision-making” (JIU 2014, iii). They are discussed both by the management and IO governing bodies and their recommendations are tracked and followed up over time. In 2018, UNDP, for instance, undertook 17 thematic and country programme evaluations (IEO 2018), while ILO conducted 54 centralized evaluations (ILO 2019). The cost of a single report can reach up to one million USD. As an example, the average cost of a thematic evaluation report in FAO is about 500,000 USD (FAO 2016, p. 28). As for the staff working in evaluation units, the numbers can vary from single figures as in the IOM to several dozen as in FAO.

The IO evaluation units are thus at the centre stage of evaluation systems in IOs. According to the UN system-wide evaluation standards, evaluation units are designed as independent actors within the IPA: “Organizational independence requires that the central evaluation function is positioned independently from management functions” (UNEG 2016, p. 11). Whereas centralized evaluation units are *formally* independent, they remain a legal entity of their IO and the IPA.

However, studies at the domestic level suggest that stakeholders can translate their influence *informally* through the institutional context in which evaluation activities are embedded (Azzam 2010; Højlund 2014; Bjørnholt and Larsen 2014). We hold that this is also true at the international level. The institutional context is the evaluation system, i.e. the organizational structure and rules that define who controls key resources and procedures of IO evaluation unit, including its budget, staff appointments, reporting lines, overall agenda, etc. (cf. Leeuw and Furubo 2008).

We therefore define *control over evaluation systems* as the independent variable (see Fig. 1). It can be exerted most directly through the evaluation staff, budget, and by means of influencing the evaluation agenda (see Rossi et al. 2004, p. 46; Stockmann et al. 2011). The core expectation is that depending on who controls the resources of evaluation system, we can observe the politics of evaluation at two subsequent stages: through the *orientation of the evaluation unit* (dependent variable 1); and at the level of *political decision-making* (dependent variable 2). We therefore specify two dependent variables as well as hypotheses linking these variables.

First, despite their formal independence, evaluation units operate as agents in an institutional setting of competing principals (Lyne et al. 2006, p. 44; Schoenefeld and Jordan 2017, p. 277), namely the two stakeholders: the IPA and member states. Given that such formal factors like funding or agenda-setting are valuable resources principals can use for sanctioning or rewarding evaluation units, there should be differences in the *evaluation unit orientation* (dependent variable 1) (Leeuw and Furubo 2008, p. 166; see also O'Brien

et al. 2010, p. 432). By evaluation unit orientation, we mean the conscious or unconscious perception by evaluation unit staff of who is the primary sponsor and user of evaluation. This does not necessarily mean that evaluators neglect professional standards and provide biased reports. Yet, according to Weiss (1998, p. 31), it should make a difference for evaluation methodology and conduct, whether the political purpose of evaluation is to generate new ideas to inform IPA policy positions or to ensure, from a member state perspective, that IPAs do not deviate from their mandate. There may also be procedural implications, such as whose comments are primarily (consciously or not) taken into account when drafting terms of references, the report, or recommendations. Overall, the expectation is that the different structural domination by either member states or the IPA determines whose political interest evaluation units will primarily orientate to. On this basis, we derive the following hypothesis on the relationship between evaluation system control and evaluation unit orientation:

H<sub>1</sub>: If the evaluation system resources are controlled by member states (or IPA) in an IO, then the evaluation unit orientates primarily towards member states (or IPA) as the sponsor and user of evaluation.

The hypothesis is falsified if we observe no systematic link between the two variables. There might also be cases with mixed stakeholder dominance, for instance, when the IPA allocates the budget, but member states approve the head of evaluation unit and evaluation agenda. In these cases, we expect evaluation units to balance between the two stakeholders without clearly orientating to either of them.

Second, depending on which stakeholder evaluation units orientate to, i.e. member states or the IPA, they should produce evaluations that tend to serve to and be used for the respective actor's political interests. The reason is, as outlined by principal-agent theory, that "the agent [here evaluation unit] scans the range of principal demands and identifies a point that maximizes the compensation offered by the various principals. Principals with more power and resources thus have a greater impact on agent behavior" (Lyne et al. 2006, p. 58). Previous research on evaluation found that "the more political power or influence stakeholder groups held over evaluation logistical factors (i.e. funding, data access), the more evaluators were willing to modify their design choices to accommodate perceived stakeholder concerns" (Azzam 2010, p. 45; see also Stockmann et al. 2011). Evaluations that serve specific political interests should also tend to be used accordingly, what we term the *evaluation use alignment* with stakeholders' interests (dependent variable 2). In line with Table 2 and depending on the value of the dependent variable 1 (and the independent variable), three patterns of political evaluation use are possible: first, in IOs with evaluation units orientated to member states, we expect evaluation to be primarily used in negotiations among member states to back certain political claims and/or evaluations used to counter unilateral state interests (collective principal dynamics). Also, we expect evaluations in such cases to be used for containing IPA's influence and avoiding bureaucratic drift (agent control). Second, in evaluation systems with units' orientation towards the IPA, we expect evaluation to be used to sanction or steer the behaviour of decentralized IPA units (collective agent) and/or evaluation use for backing or justifying the IPA's own initiatives (bureaucratic influence). The second hypothesis below summarizes our expectations (see also Table 2):

H<sub>2</sub>: If evaluation units are orientated to member states (or IPA) as their primary sponsor and evaluation user, evaluation use aligns with typical member state (or IPA) political interests.



The hypothesis is falsified if we observe no systematic link between the evaluation unit orientation and the political use of evaluation results. Again, in cases of mixed evaluation unit orientation, we expect competition between stakeholders and thus mixed political evaluation use aligned to the interests of both member states and the IPA.

## Research design

The cause–effect relationship outlined by the two hypotheses implies a causal mechanism, which scholars often test through in-depth case studies along a temporal dimension (Beach and Pedersen 2016; Mayntz 2004). Yet, contrary to a processual approach, which restricts the number of cases that can be studied and thus impairs the external validity of the findings, we apply a comparative most similar systems design (MSSD) according to Mill’s method of difference (Lijphart 1971). Medium-N controlled comparison allows testing theorized claims for a broader set of cases and provides for generalization beyond the selected cases (George and Bennett 2005).

Case selection proceeded in two steps. First, we selected nine IOs which are similar on alternative explanatory dimensions but vary in the independent variable (control of evaluation system resources). IOs in the sample are: the IAEA, IOM, ILO, UNICEF, UNDP, UNEP, UNHCR, UNESCO, and WHO. They all belong to the UN system and as such have a similar membership with governing bodies consisting of member states representatives and a relatively similar IPA structure oriented towards common UN staff policies. As they are all UN organizations, their evaluation activities are defined around the same guidelines, norms, and standards of the UN Evaluation Group. They all have institutionalized evaluation units which are responsible for the centralized evaluation function according to their evaluation policies.<sup>5</sup> They all conduct evaluations of the IPA’s activities and present annual reports both to the senior management and member states. Finally, evaluation units operate independently from management in formal terms.

Despite these similarities, the nine IOs differ in who formally controls evaluation system resources. Accordingly, IOs were grouped into three clusters. Cluster 1 contains ILO, UNDP, UNICEF whose evaluation system resources are predominantly controlled by member states; cluster 2 includes IAEA, IOM, and UNHCR, where the head of the IPA—and not the member states—decides upon the evaluation system resources; and cluster 3 consists of UNEP, UNESCO, and WHO. In these IOs, none of the resources is approved unilaterally by member states or the IPA.

The operationalization of evaluation system control (independent variable) was based on the following procedure. We focused on evaluation staff, budget, and agenda as these are the key evaluation resources identified in the literature (Azzam 2010; Stockmann et al. 2011; Rossi et al. 2004). Drawing on official evaluation policies of all IOs (see “Appendix 1”), we used a scoring system by allocating 1, 0.5, or 0 points to either member states or the IPA. 1.0 point was allocated to the IPA or member states if they unilaterally approved evaluation budget, agenda, or appointment of the evaluation unit head. For instance, according to the UNHCR Policy on Evaluation, the High Commissioner (head of IPA) approves “the annual Work Plan for centralised evaluations and the Evaluation Service

---

<sup>5</sup> These units may conduct evaluations completely by themselves or recruit external consultants. Regardless of which business model is chosen, final reports are under the authority of IO evaluation unit, which makes the distinction between in/house and external consultants less relevant.

**Table 3** Operationalization of evaluation system control and coding rules

Resource	Coding rule
Stakeholder	MS
Code	1
	IPA/MS (both)
	0.5
	1
<b>Evaluation budget</b>	Member states (IO governing body) approve the evaluation budget
	The head of IO administration approves evaluation budget
<b>Evaluation agenda-setting</b>	Member states approve the rolling work plan of evaluation unit
	The head of IO administration approves the rolling work plan of evaluation unit
<b>Evaluation staff appointment</b>	Member states (IO governing body) appoint the head of evaluation unit
	The head of IO administration appoints the head of evaluation unit

**Table 4** Expected values to test hypotheses 1 and 2 across 6 IOs clustered in three groups

IO sample	Anonymized label	Control of evaluation resources (iV)	Expected evaluation unit orientation (dV1)	Expected evaluation use alignment (dV2)
ILO, UNDP, UNICEF	Cluster 1 (IO 1, IO 2)	MS	MS	MS
IAEA, IOM, UNHCR	Cluster 2 (IO 3, IO 4)	IPA	IPA	IPA
UNEP, WHO, UNESCO	Cluster 3 (IO 5, IO 6)	MS/IPA	MS/IPA	MS/IPA

Budget” (UNHCR 2016, p. 12). By contrast, in UNDP, the Board “approves the biennial financial appropriation to the Independent Evaluation Office” as well as “the programme of work of the Office” (UNDP 2016, p. 6). 0.5 point was allocated to both stakeholders if they shared the authority to approve these resources. For example, according to WHO Evaluation Policy, the head of evaluation unit is appointed by the Director-General, yet “after consultation with the Executive Board” (WHO 2018). Table 3 summarizes the coding rules.

If one of the stakeholders scored 2.5 or 3 points, the IO was attributed to either cluster 1 (MS control cases) or cluster 2 (IPA control cases) accordingly. If both stakeholders scored more than 0.5 points, the IO was attributed to mixed control cases of cluster 3. See “Appendix 1” for specific references from IOs’ evaluation policies on each resource category and resulting mapping of evaluation system resources.

In the second step of case selection, we randomly selected two out of three IOs per cluster and gathered empirical data on these six IOs. The reason being, we sought to ensure the anonymity of interviewees: the number of evaluation staff per IO is usually low and the topic of this study is sensitive to respondents’ professional integrity. Previous studies on similarly contentious topics also anonymized IOs under scrutiny (e.g. Mele et al. 2016). Hence, from this point on, we no longer reveal the identity of each IO but refer only to the clusters (see Table 4).

The selected six UN organizations are similar on alternative factors that might confound evaluation processes as outlined above. In line with Mill’s method of difference, the observation of the variable values as indicated in Table 3 should allow conclusions on the hypotheses. A discussion on alternative explanations is still provided below.

To measure *evaluation unit orientation* (dV1) and *evaluation use alignment* (dV2, we draw on original expert interview data collected for this purpose. We conducted 35 semi-structured interviews with officials from three target groups in 2018 and 2019: heads of IO evaluation units, member state representatives (mostly ambassadors) from those governing bodies to which evaluation units report (e.g. executive board, programme committee, executive committee), and senior management officials from respective IOs (e.g. Chef de Cabinet or programme directors). We made sure that all stakeholder groups (evaluators, member states, and IPA) are equally covered (see interview list in “Appendix 2”).

To avoid biasing interviewee responses, the questionnaire contained only general questions about the evaluation function in an IO (its system, process, use; see “Appendix 3”). We did not directly ask about orientation and political use of evaluation results. While evaluators or management officials might be aware of the political nature of evaluation in their organization, it is likely that not all would have accurately responded to such questions. Instead, we measure the key variables indirectly, by assessing what meaning interviewees

bring to evaluation and its conduct. This is in line with how Denzin and Lincoln (2000, p. 3) describe the approach: “qualitative researchers study things in their natural settings, attempting to make sense of, or to interpret, phenomena in terms of the meanings people bring to them”. All interviews, each lasting about 60 to 90 min, were recorded and transcribed.

To measure the two dependent variables, we systematically coded interview transcripts using MAXQDA 2018 software. *Evaluation unit orientation* (dV1) was operationalized using two codes. First, all text segments in which evaluation unit respondents referred to member states or IPA as their sponsors (i.e. those who help to fulfil their mandate) or identified themselves with these actors as being part of them were given a code “Evaluation Unit Sponsor”. Depending on whom they referred to, directional sub-codes were attached (either MS or IPA). Second, the same procedure was conducted for text segments where evaluation unit respondents referred to either member states or the IPA as the main users of evaluation results (“Evaluation Unit User”).

Analogously, we followed the same approach to measure *evaluation use alignment* (dV2). Text segments from all interviews where respondents made statements about evaluation use were classified according to the indicators from Table 2: directional codes indicated evaluation use alignment either towards member states’ or an IPA’s political interests (e.g. “MS use/collective principal” or “IPA use/bureaucratic influence”).<sup>6</sup>

In total, 236 text segments (i.e. interviewees’ statements on evaluation unit orientation or political evaluation use) were coded. The interviews were coded by two researchers, focusing on text segments identified as relevant by the first author, and then by resolving any disagreements (see “Appendix 4” for the overview of coded statements per IO and stakeholder group).

Note that our results do not account for variation over time and thus refer to the situation as of 2018/19. Furthermore, we were only able to interview a limited number of stakeholders, which requires us to generalize on the interests of member states or the administration based on statements made by a small number of their representatives. This is, however, a general issue in qualitative studies. Nevertheless, we sought to enhance the representativeness of our interviewees by reaching out primarily to longstanding member state representatives and high-level IPA officials who have a broader view on the overall interest dynamics among their fellow members. Finally, we do not claim to cover specific particularities of individual evaluation processes; our aim is to reveal general patterns of evaluation use across a number of IOs.

## Results

In the following sections, by going from IO cluster 1 to IO cluster 3, we navigate through the two dependent variables and their values for each IO and provide illustrations of the theorized dynamics. The interviews are labelled according to the target group (E for evaluators, MS for member states, IPA for secretariat staff). Note that we also report the share of directional sub-codes for each organization, which serves as an additional indicator for variable values. The aim is not to artificially quantify qualitative data. Instead, we aspire

---

<sup>6</sup> Naturally, interviewees also made generic references to learning and accountability. Given that we are interested in the evaluation use along political interests linked to P-A-dynamics, we do not report them.

transparency and reliability as to how we decided about the dominant pattern in each IO. It also helps to summarize and illustrate findings.

### **Variation of evaluation units' orientation (dV1)**

*Cluster 1* (selected from ILO, UNDP, UNICEF) includes two IOs where member states control both the evaluation budget and agenda. In line with hypothesis 1, we found that in both IO 1 and IO 2 evaluation units tend to clearly orientate themselves towards the member states as their primary principals.

In IO 1, the head of evaluation unit regularly noted that the policy direction “is not set by the bureaucrats, it is set by the member states” and described their interest in evaluation as the “most important question”: one can only make evaluations “as useful as possible by ensuring that you address issues that your constituent and policy-makers are interested in” (E1). This perfectly illustrates what we mean by evaluation unit’s orientation towards one of the stakeholders. Furthermore, the evaluation unit’s director claimed to have the mandate from the governing body to force the IPA to follow his or her requirements, stressing that “if I go back to the governing body and say that there has not been any follow up on the evaluations, there is no one that can stop me”. Such a protection from the member states was implied to be crucial: “unless I have that, I would not be able to hold back the pressure [from the management]” (E1). By contrast, the respondent referred to the IPA mostly in the context of technical issues.

Similarly, in IO 2, the head of evaluation unit highlighted several times that he or she works independently from the administration which is not the evaluation unit’s client (E2). Instead, the interviewee highlighted consultations with member states as a “critical process” to evaluation and explained that the evaluators’ task was to evaluate—not to aid—the administration (E2). While this sometimes created “a bit of a shock” for the administration, it brought “respect at the [Executive] Board” (E2). The orientation towards member states was illustrated with several examples, where member states supported the evaluation unit vis-à-vis the management: for instance, when IPA staff requested content alterations in evaluation reports, refused to give access to information, or refrained from providing a management response. The interviewee even revealed that such conflictive situations led to attempts by the IPA to alter evaluation policy and intervene into the evaluation unit’s independence. However, due to the support by the member states, such attempts were “lost badly” (E2).

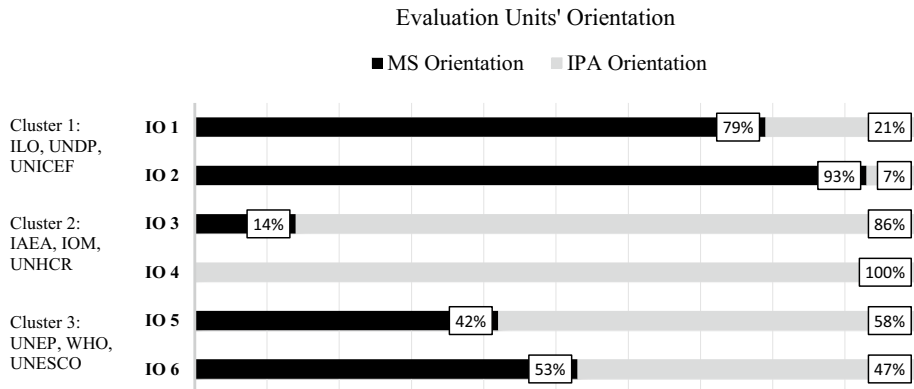
*Cluster 2* (selected from IAEA, IOM, UNHCR) includes 2 IOs where the IPA controls both the evaluation budget and agenda. Our empirical data from IO 3 and 4 clearly demonstrate evaluation units’ orientation towards the IPA as a key sponsor and user, which, again, supports hypothesis 1. The head of IO 3 evaluation unit claimed that the IPA and not member states was the key beneficiary of evaluation service. It was stated that “member states get what they want, but they get it not through the evaluation service” (E3). On the one hand, the interviewee stressed the unit’s direct reporting to the head of the administration, while on the other it was noted that member states’ involvement in evaluation activities is rather low and fragmented. This, in turn, shows that the target group of the evaluation unit was the IPA, rather than the governing body. As the head of the unit said, draft results were shared with “the teams that are most associated with the evaluated work” and (s)he would do briefings or “a workshop where we are talking about the findings and the recommendations” (E3).

The evaluation unit of IO 4 also said that the IPA was the key target group for all their evaluations. The interviewee even claimed that it is in agreement with the member states, as they “themselves want to give certain freedom to the Director-General to manage the organization” (E4). Furthermore, although only few statements were made on the indicator of “Evaluation Unit Sponsor”, the interviewee associated him- or herself with the bureaucracy, highlighting the active and flexible internal interaction at different levels, where people know each other and operate “as kind of a family” (E4).

Finally, *cluster 3* (selected from UNEP, WHO, UNESCO) includes IOs where evaluation systems are dominated by both member states and the IPA. Their heads depicted both member states and the IPA as their main sponsors and evaluation users. In IO 5, the unit head emphasized the importance of member states’ push towards a stronger evaluation function and stressed their recognition of good evaluators’ work (E5). It was argued that “member states assign more and more tasks” and that evaluations are often done “because member states wanted us to do it” (E5). Yet, on the other hand, the official regularly referred to the mutual goals with the rest of the IPA and how evaluation helps the organization to improve. Various formats of “internal” discussions within the organization were outlined before proposals would go to member states (E5). Finally, a mixed message regarding the unit’s sense of sponsorship was given. The interviewee explained that the IPA’s head gave him or her the mandate “to go ahead” and ask everyone to “open everything [he or she] needed to see”. At the same time, it was mentioned that member states watch the evaluation work “with a kind of covering hand” and demand it to be independent (E5).

In IO 6, when describing his or her function, the evaluation head claimed to “provide timely advice” to the head of the IPA, implying the administration to be the primary user (E6). However, it was also mentioned that “at the end of the day, we are a member states organization and if they have capacity, interest, and majority to run a certain thing..., there is nothing we can do about it” (E6). Regarding sponsorship, the interviewee said (s)he could openly discuss all matters with the senior management *before* the discussion with the governing body, which seems to be a relevant opportunity for the evaluation unit to ensure management support given that member states tend to pressure the evaluation director to answer very “pedantic” questions. However, further statements were also made on member states’ sponsorship in terms of political protection from management’s influence saying that “nobody can stop us from writing and sending things to the [Executive] Board... This helps us to build credibility with the member states” (E6). Such contradictions thus indicate a mixed evaluation unit’s orientation to both member states and the IPA.

Figure 2 summarizes our findings and illustrates the dominant pattern in the six IOs based on our coding results. A clear trend towards the member states as primary principals can be observed in the responses of interviewees from IO 1 and IO 2, where the overwhelming majority of statements made on evaluation unit orientation were directed to member states (79–93%). In contrast, interviewees from cluster 2 IOs mostly (or even exclusively for IO 4) spoke about the IO administration as a key sponsor and user of evaluation. In IO 5 and 6, statements referred both to the IPA and IO member states, showing a mixed pattern in evaluation units’ orientation. All three IO clusters tend to maintain the first hypothesis, allowing us to conclude that evaluations units orientate themselves towards actors who controls the structure of evaluation system. The next question is whether we can also observe a corresponding pattern in the political use of evaluation results.



**Fig. 2** Illustration of dominant patterns in evaluation unit orientation across six IOs (dV1). The results indicate what percentage of all statements on evaluation unit orientation (made by evaluation unit interviewees) were directed either to member states or the IPA

### Variation of political evaluation use alignment (dV2)

Analogously to the previous section, we now turn to the empirical analysis of our second hypothesis. Other than in the previous section, we now include interviews with member state (MS) and administrative representatives (IPA). We report answers corresponding with evaluation use along member states and IPA interest as identified in the theory section (we refer to the dimensions in Table 2 with *terms in italics*).

The findings from *cluster 1* (selected from ILO, UNDP, UNICEF) reveal a pattern of evaluation use predominantly aligned to the interests of member states. In IO 1, when asked about evaluation use, most respondents highlighted the complexity of member states with “conflicting interests among parties”, where evaluation helps to learn about each other’s preferences (MS1; E1) (*collective principal*).<sup>7</sup> Interviewees also claimed that member states would impose evaluation recommendations as a tool of control if the IPA would not go “in line with the policy direction that they [member states] have in mind” (E1; MS1; IPA1) (*P-A control*). By contrast, only very few statements were made on typical IPA interests. An IPA official said that the administration may occasionally use results as “a negotiation card” within the IPA’s own structure for distinct departments to get support from the senior management (IPA1).

In IO 2, respondents also highlighted member states’ heterogeneity, arguing that evaluation is often used in negotiations between states (IPA3; IPA4). Such use was clearly political—due to their “national views”, states would only focus on evaluation findings or topics that suit their purposes, making evaluation a “contested area” (MS2; MS5; MS6). As one IPA official explained, such contestation was especially tangible between developing and developed countries (IPA4) (*collective principal*). Furthermore, interviewees referred to the need to contain the IPA. Member states expressed their concerns that the IO management downplays issues and shows “the good things” only (MS3, MS2). Others suggested that the bureaucracy has its own agenda that is hidden from member states (MS4). In light of this, evaluation was depicted as a tool to control the administration, i.e. to get “an outside view”

<sup>7</sup> If several interviews are cited, the first label refers to the direct quotation, whereas further interviews refer to similar statements.

(MS4), to reveal “where the challenges are” (MS3), and to “send signals” on course correction (MS2). The IPA staff, too, described evaluation as a “top down external mechanism” which member states use to “control and to hold you [IPA] to task” (IPA5) (*P-A control*).

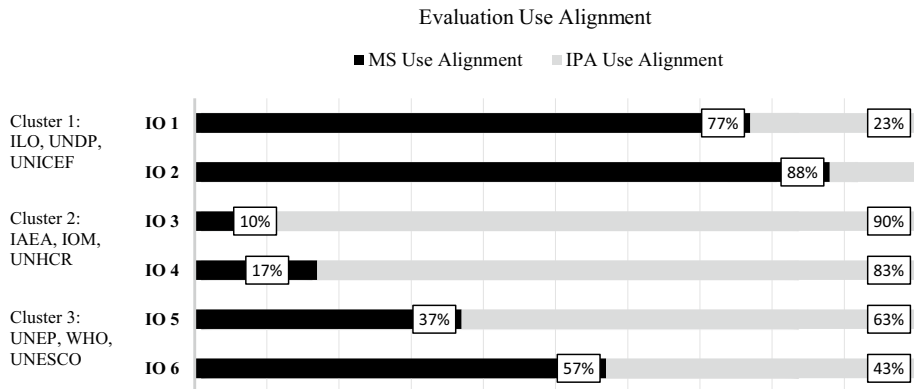
In contrast, *cluster 2* (selected from IAEA, IOM, UNHCR) indicates a strong tendency towards evaluation use aligned to IPAs’ political interests both regarding internal fragmentation dynamics and in relation to member states. In IO 3, evaluation was described as a tool to gather information on key issues at all levels of organization, from junior staff to senior management, which then allows the administration leadership to set strategic priorities and internally steer the organization by “creating course corrections” (E3; IPA8; MS7; MS8) (*collective agent*). Interviewees also said that the senior management might manipulate member states’ positions on certain issues (e.g. country programs) by framing evaluation results “in a politically clever or politically sensitive way” (MS9; IPA7; IPA8). Such a tactic was described as “advocacy on issues” in a dialogue with governments (E3), which perfectly illustrates how evaluation may help the IPA to gain policy-making influence. It was also noted that the IPA may use evaluations to justify performance failures, to show “that things are going better and that it is in their [member states] favor” (IPA7; MS7) (*bureaucratic influence*).

A similar pattern prevails for IO 4. Regarding vertical IO steering, evaluation was said to be helpful to consolidate internal “information and knowledge management” (MS9; E4; IPA10), referring to the IPAs fragmented organizational structures (*collective agent*). In relation to IPA influence and justification, interviewees claimed that evaluation helps the organization to raise additional funding (to “sell projects”) (MS8) and allows the IPA to “promote” its own activities or “justify own mistakes”; for instance, by drawing member states’ attention to specific evaluation reports in governing body meetings (E4; IPA9). It was even claimed that evaluation reports would never admit the IPA’s own mistakes but would rather point out the external factors (IPA9) (*bureaucratic influence*). On rare occasions, respondents also noticed that member states would refer to evaluation to oversee the IPA or convince other member states, for instance, in budgetary questions (IPA10).

Finally, a more mixed pattern prevails in *cluster 3* (selected from UNEP, WHO, UNESCO). In IO 5, interviewees again described the diversity of member state interests and claimed that the organization shall “deliver benefit for all the membership and not only for some” (MS11; MS1; MS10; E5) (*collective principal*). Furthermore, evaluation was argued to “detect” issues which might not be reported by the secretariat itself (*P-A control*) (MS11; MS8; IPA12; IPA13). But a significant number of interviewees also pointed out that evaluation served typical IPA interests. For instance, interviewees claimed that evaluation often helps senior management to get a joint understanding about “the different departments doing their own thing” (*collective agent*) (IPA12; IPA14; E5). Interviewees also detailed that evaluation was strategically used by the head of IPA to ask member states for additional funding (MS11), to promote specific programs (MS1), or convince member states if needed: “It makes it much easier to convince people because it is... evidence based. Using evaluation in communication parts is extremely powerful” (*bureaucratic influence*) (IPA14).

In IO 6, interviewees talked about “political struggles” among member states and how evaluation may help to counter single countries trying “to defend their programs... regardless of the results” (*collective principal*) (MS12; MS13; E6). Interviewees also revealed that “the administration is trying to keep the member states out...” (IPA15), whereas evaluation findings provide member states with a baseline for comparison of certain programs (MS12; MS13; IPA16) (*P-A control*). But again, almost half of all statements referred to distinctive IPA interests. For instance, the IO leadership said that they benefited from getting “insight into the processes and the functioning of individual offices where we [senior





**Fig. 3** Illustration of dominant patterns in political evaluation use alignment (dV2). The results indicate what percentage of all statements on political evaluation use (made by all interviewees) were directed either to the interests of the member states or the IPA

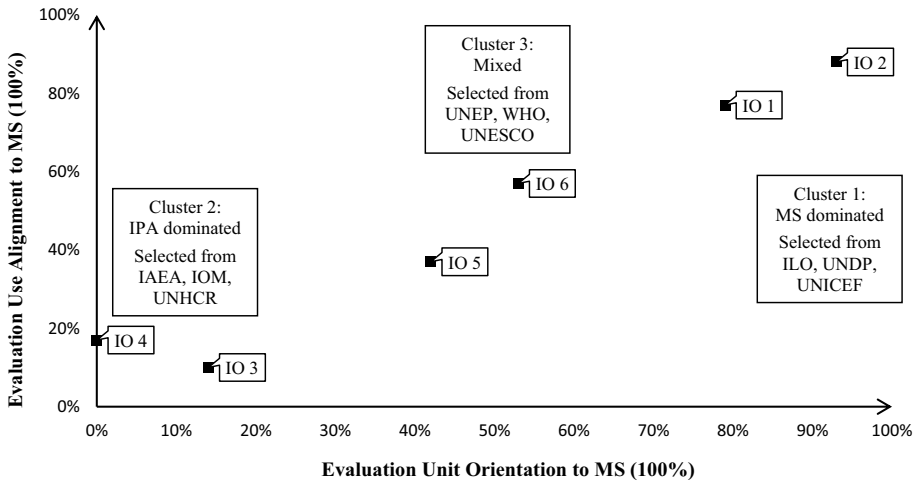
management] do not have sufficient view on what is going on” (IPA15; IPA16) (*collective agent*). Finally, it was also argued that the IPA would “regularly refer to evaluation reports to defend its own position” *vis-à-vis* the member states, especially when it comes to budgeting or the extension of projects or programs (MS13) (*bureaucratic influence*).

Figure 3 summarizes the findings. The share of interviewee statements referring to either political use alignment with IPA interests, or use alignment with member state interests (or the mix thereof), corresponds with the theorized assumption that the evaluation unit’s orientation determines the type of political evaluation use. Admittedly, evaluation use alignment towards IPA in IO 5 is more pronounced than we expected (with 63%); however, the pattern is still clearly balanced if taken together with IO 6 and compared to the other two clusters.

## Discussion

Overall, the tendency of interviewees—across all three target groups—to frame their statements in the theoretically expected direction is remarkably strong. We find all expectations well confirmed (see summary in Fig. 4). In line with  $H_1$ , the evaluation units’ orientations were consistent with our expectation that those who control evaluation resources will be perceived as primary sponsors and evaluation users. In line with  $H_2$ , we found that political evaluation use followed the expected pattern. In cluster 1 (member state dominated IOs), respondent statements on political evaluation use referred to typical member states’ interests (containing unilateral influence; controlling the IPA). In cluster 2 (IPA dominated IOs), respondents predominantly mentioned typical IPA political interests (justification, policy influence and internal steering). In IOs with balanced systems (cluster 3), evaluation unit orientation and use alignment are also mixed.

The main limitation of this study is the lack of data at the level of evaluation reports and at the level of policy outcomes as our data on political evaluation use consists of statements made by experts who described the use of evaluation results in policy-making processes of their IO in general terms. At the same time, the strength of our approach is that we did not ask about the political use of evaluation directly but measured both dependent variables indirectly by comparing how stakeholders responded differently to similar general questions on evaluation use, depending on who controls evaluation system resources. While



**Fig. 4** Illustration of empirical findings based on coding results. The results indicate the share of total interviewee references to evaluation unit orientation and evaluation use alignment with 100% meaning that all of the coded statements in the category referred to member states

findings therefore remain generic, we expect nonetheless that the dynamics theorized and described in this paper should have substantive implications, both for actual evaluation research and reports (see for instance the results by van Voorst and Mastenbroek 2019) and for policy decisions (see also the substantive evidence about pressure on evaluators; (cf. Pleger et al. 2017). Whereas other research designs are necessary to answer detailed questions about policy outcomes, the empirical merit of this paper is in-depth insights from main evaluation stakeholders, particularly evaluation unit experts.

Finally, there is little reason to believe that confounding factors drive our results. On the one hand, the six IOs studied in this paper subscribe to the same evaluation norms and standards, conduct centralized evaluations and report to similar governance structures, which rules out alternative explanations *linked to evaluation policy*. On the other hand, the IOs vary in their mandate, policy field, and operational profile both within and across the clusters (see “Appendix 5”) This rules out alternative explanations *linked to general IO characteristics*. It is reasonable to conclude, therefore, that evaluation unit orientation and political evaluation use in other comparable IOs should also vary according to the theorized patterns.

## Conclusion and theoretical implications

In this paper, we propose a relatively simple, yet novel, theory-founded framework on how to study the political use of evaluation. This is especially useful in times of booming evaluation businesses and the proliferation of institutional assessments both in domestic and international politics (see Cooley and Snyder 2015). Although evaluation is still “widely viewed as the ‘gold standard’ of institutional assessment” (Lall 2017, p. 245), we demonstrate that evaluation systematically serves ex-ante political interests of policy actors, depending on who exerts control over evaluation system resources.

Our findings yield two theoretical implications. The first speaks to the literature on IOs and International Relations. The findings show that institutional IO design matters. At the same time, our results disfavour the *state*-centric view of international institutions which is embedded in rational design theory (Abbott and Snidal 1998; Koremenos et al. 2001). As our findings from cluster 2 and 3 demonstrate, IO *bureaucracies* are able to use IOs' internal structure (evaluation system) to exert their influence using evaluation as a political tool (Patz and Goetz 2019; Knill et al. 2019; Biermann and Siebenhüner 2009; Bauer et al. 2017; Johnson and Urpelainen 2014). The key theoretical puzzle for future research is to explain the *reasons* for such variation in control over evaluation resources.

Secondly, our results speak to the literature on Public Policy and Evaluation. The findings underline that the purpose, efficiency, and function of public management tools, including evaluation, should always be seen in the context of contested stakeholder interests, especially in such complex organizational environments like IOs. For anthropologists, the expansion of evidence instruments, indicators, and quantification “comes from a political culture that demands more openness and seeks to drive out corruption, prejudice, and the arbitrary power of elites” (Merry 2011, 85; see also Porter 1995). The paper's findings, however, raise doubts that such a culture can ever be realized in the context of public service organizations. While IOs may be extreme in their internal political contestation (see our literature review), bureaucratic politics and the dichotomy of administration and political actors are also well known domestically. Neglecting the political nature of evaluation stimulates unintended consequences, when functional tools are harbouring political agendas. In this regard, studies on evaluation should generally pay more attention to the setup of evaluation systems, who controls evaluation resources and how such differences came to be (see Hinterleitner et al. 2016; Fforde 2019; Perl et al. 2018, p. 591; van Voorst and Mastenbroek 2019).

In terms of practical implications, practitioners should consider that political use of evaluation may hinder its functional purposes. As our data implies, an evaluation unit's strong orientation towards one stakeholder leads to disengagement by the other stakeholder. For instance, in IO 2, where member states dominate the evaluation system, IPA officials were sceptical about evaluation's contribution to learning. They criticized the quality of evaluation reports and explained that evaluation is always “at the center of controversy”, triggering “defensive behaviors” among staff (IPA6, IPA5). By contrast, in IO 3, where the evaluation unit is controlled by the IPA, member states were less interested in evaluation's benefits and perceived it as an internal “management tool” (MS7).

Future research should thus investigate whether evaluation systems with the mixed control setting (cluster 3 IOs) mitigate politicization and increase evaluation's functional use. After all, the proper answer to evaluation politics is to improve, and not abandon, the evaluation practices. Scholars should also examine the extent to which political interests affect the actual evaluation research and results. Today, evaluation has become a booming industry and it should be relevant whether we are merely looking at politics with other means.

**Funding** Funding was provided by Deutsche Forschungsgemeinschaft (Grant No. EC 506/1-1). Open Access funding provided by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the

material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix 1: Measurement of stakeholder control over evaluation systems

### Primary data sources

IAEA	IAEA Evaluation Policy (2011), Office of Internal Oversight Services (OIOS) OIOS Charter (2014)
ILO	ILO Evaluation Policy (2017), GB.331/PFA/8 ILO policy guidelines for evaluation: Principles, rationale, planning and managing for evaluations (2017), 3rd edition
IOM	IOM Evaluation Policy (2018) Charter of the Office of the Inspector General (OIG) (2015), IN/74 Rev.1 IOM Evaluation Guidelines (2006), OIG
UNICEF	Revised Evaluation Policy of UNICEF (2018)
UNDP	Revised UNDP Evaluation Policy (2019)
UNEP	UNEP Evaluation Policy (2016), Evaluation Office
UNHCR	UNHCR Policy on Evaluation (2016), UNHCR/HCP/2016/2 UNHCR Evaluation Strategy (2018–2022)
UNESCO	UNESCO Administrative Manual (2017), Item 1.6 Internal Oversight UNESCO Evaluation Policy (2014–2021), IOS/EVS/PI/162
WHO	Evaluation: evaluation policy (2018), EB143/9

These documents were used to measure the control over evaluation system resources (independent variable), see below

### Measurement of stakeholder control over evaluation system resources

IO	Evaluation staff appointment	Evaluation budget allocation	Evaluation agenda-setting
ILO <i>Source:</i> 2017 ILO Evaluation Policy	MS&IPA 0.5 Not specified in the policy, but: "In terms of our reporting line, I report only to the DG and the governing body" (interview)	MS 1 The budget approval procedure not specified, but: "The governing body is the decision-making body of the ILO. They have to proof the budget..."; "Budget of EVAL Office decided by constituents in government body meetings" (interview)	MS 1 "EVAL will propose to the Governing Body each year... a proposed rolling programme of evaluation work for major independent evaluations"

IO	Evaluation staff appointment	Evaluation budget allocation	Evaluation agenda-setting
UNDP <i>Source:</i> 2019 UNDP Revised Evaluation Policy	IPA/MS 0.5 “The UNDP Administrator appoints the Director of the Office in consultation with the Executive Board, taking into account the advice of the Audit and Evaluation Advisory Committee”	MS 1 “The Board approves the biennial financial appropriation to the Independent Evaluation Office”	MS 1 “The Board... undertakes periodic reviews and adjustments of such appropriations based on the programme of work of the Office, which the Board also approves”
UNICEF <i>Source:</i> 2018 Revised Evaluation Policy	MS&IPA 0.5 “The Director of Evaluation is appointed by the Executive Director in consultation with the Audit Advisory Committee and the Executive Board, with an external evaluation expert as part of the selection panel”	MS 1 “As part of the approval of the integrated budget, the Board approves the budget of the Evaluation Office”	MS 1 “The Executive Board... approves the plan for global evaluations”
IAEA <i>Source:</i> 2014 IAEA OIOS Charter	IPA 1 Evaluation policy does not specify appointment rules, but: “the Director of OIOS... reports directly to the Director General”	IPA 1 “The annual work plan is prepared and proposed to the Director General to ensure that the resources required for the internal oversight services are sufficient, appropriate, and effectively deployed to meet the expected outcomes”	IPA 1 “The annual (OIOS) work plan shall be subject to the approval of the Director General”
IOM <i>Source:</i> 2015 Charter of the Office of the Inspector General (OIG)	IPA 1 The head of evaluation unit is appointed by the Inspector General, who “is appointed by the Director General, who shall consult with the AOAC [advisory committee for the DG]”	IPA 1 “The Director General ensures that OIG is provided with the necessary resources in terms of ...adequate funds...to fulfil its mission and maintain its independence”	IPA 1 Biennial evaluation plan is submitted to the Director-General through the Audit and Oversight Advisory Committee (AOAC)
UNHCR <i>Source:</i> 2016 UNHCR Policy on Evaluation	IPA 1 “The High Commissioner is responsible for...appointing the Head of the Evaluation Service with the required experience, expertise, profile and qualifications”	IPA 1 “The High Commissioner is responsible for...with support by the Deputy High Commissioner, approving the annual Work Plan for centralised evaluations”	IPA 1 “The High Commissioner is responsible for...with support by the Deputy High Commissioner, approving...the Evaluation Service Budget”

IO	Evaluation staff appointment	Evaluation budget allocation	Evaluation agenda-setting
UNEP <i>Source:</i> 2016 UNEP Evaluation Policy	IPA 1 “The Executive Director ensures that adequate and qualified staff are recruited for the effective functioning of the EOU”	MS 1 “The UNEA approves the operational budget of the EOU by reviewing and approving the EOU’s allocation within the overall budget of the organization of UNEP’s PoW”	IPA 1 “The biennial evaluation work plan will be reviewed by the SMT and approved by the Executive Director”
WHO <i>Source:</i> WHO Evaluation Policy 2018	MS&IPA 0.5 “The Director-General shall appoint a technically qualified head of the Evaluation Office after consultation with the Executive Board. The Director-General shall likewise consult the Executive Board before any termination of the incumbent of that office”	MS 1 “The Executive Board shall approve the biennial Organization-wide evaluation workplan, including its budget “	MS&IPA 0.5 “The workplans shall be submitted to the Executive Board for approval through the Programme, Budget and Administration Committee” BUT: The workplan will be finalized by including mandatory evaluations and those “significant for the organization and the management” (interview)
UNESCO <i>Sources:</i> 2015 UNESCO Evaluation Policy Internal Oversight Charter- UNESCO	MS&IPA 0.5 “The Director of IOS is appointed by the Director-General. The Director-General shall take decisions concerning the appointment, extension, renewal and termination of appointment of the Director of IOS in consultation with the Executive Board (Staff Regulations and Staff Rules, Regulation 4.5.3)”	MS&IPA 0.5 “The Director-General ensures that adequate resources are allocated to implement the quadrennial evaluation plan” AND “Executive Board ensures that adequate resources are allocated to implement the quadrennial corporate evaluation plan”	MS&IPA 0.5 “The Evaluation Office establishes the quadrennial plan in consultation with UNESCO senior management... The Executive Board may also request that specific topics be included in the evaluation plan”. “The IOS Director maintains the ultimate authority for approving or modifying the corporate evaluation plan”

## Appendix 2: Interviewee list and positions

### Interview list

No.	Interview label	Organization label	Target group	Interviewee position	Date
1	E1	IO 1	Evaluation unit	Director	June 2018
2	E2	IO 2	Evaluation Unit	Director	March 2019
3	E3	IO 3	Evaluation Unit	Director	June 2018

No.	Interview label	Organization label	Target group	Interviewee position	Date
4	E4	IO 4	Evaluation Unit	Director	June 2018
5	E5	IO 5	Evaluation Unit	Director	June 2018
6	E6	IO 6	Evaluation Unit	Director	July 2018
7	IPA1	IO 1	IPA	Senior Programme Officer	October 2018
8	IPA2	IO 1	IPA	Programming Unit Head	April 2019
9	IPA3	IO 2	IPA	Unit Head	February 2019
10	IPA4	IO 2	IPA	Unit Head	February 2019
11	IPA5	IO 2	IPA	Senior Management Officer	March 2019
12	IPA6	IO 2	IPA	Senior Advisor	March 2019
13	IPA7**	IO 3	IPA	Senior Advisor	June 2018
14	IPA8	IO 3	IPA	Division Director	January 2019
15	IPA9	IO 4	IPA	Country Office Head	October 2018
16	IPA10	IO 4	IPA	Senior Manager	November 2018
17	IPA11	IO 4	IPA	Programme Head	October 2018
18	IPA12	IO 5	IPA	Department Director	June 2018
19	IPA13	IO 5	IPA	Department Director	June 2018
20	IPA14	IO 5	IPA	Senior Management Officer	June 2018
21	IPA15	IO 6	IPA	Division Director	July 2018
22	IPA16	IO 6	IPA	Senior Advisor	April 2019
23	MS1	IO 1, 3, 4, 5*	MS	Division Head	June 2018
24	MS2	IO 2	MS	First Secretary	February 2019
25	MS3	IO 2	MS	Deputy Permanent Representative	February 2019
26	MS4	IO 2	MS	First Secretary	February 2019
27	MS5	IO 2	MS	Minister	February 2019
28	MS6	IO 2	MS	Senior Advisor to Ambassador	February 2019
29	MS7	IO 3	MS	First Secretary	January 2019
30	MS8	IO 1, 3, 4, 5*	MS	Ambassador	June 2018
31	MS9	IO 1, 3, 4, 5*	MS	First Secretary	June 2018
32	MS10**	IO 1, 3, 4, 5*	MS	Minister Counsellor	June 2018
33	MS11	IO 5	MS	First Secretary	June 2018
34	MS12**	IO 6	MS	Ambassador	July 2018
35	MS13	IO 6	MS	Deputy Permanent Representative	July 2018

\*The interviewed country representation covers more than one IO from our sample, e.g. in Geneva

\*\*Upon request of the interviewee, this interview was not recorded. Instead, we took notes throughout the interview and uploaded our script into MAXQDA

### Equal levels of position seniority covered across IOs

Please note that we conducted interviews with officials from three target groups. First, we spoke with heads of each IO's evaluation unit. Thus, for evaluators, the level of seniority is held constant.

Second, we interviewed member state representatives focusing on high-level officials (ambassadors, deputies, minister counsellors) or medium-level officials (senior advisors,

division heads) as they have the best overview of evaluation-related dynamics in respective IOs' governing bodies. In some cases, lower-level diplomats were interviewed as well, for instance, when they were specifically responsible for evaluation-related issues in an IO. As the table below demonstrates, all IOs are covered equally in terms of interviewees' position levels (sorted based on interview list above). The only exception is IO 6 where we did not get access to medium-level officials. However, this should not bias the results as officials from high-level positions are covered.

IO	Interviews with member state representatives
IO 1	High level (ambassador, minister counsellor) Medium (division head) Lower (first secretary)
IO 2	High level (dep. permanent representative, minister) Medium (senior advisor to ambassador) Lower (2× first secretary)
IO 3	High level (ambassador, minister counsellor) Medium (division head) Lower (2× first secretary)
IO 4	High level (ambassador, minister counsellor) Medium (division head) Lower (first secretary)
IO 5	High level (ambassador, minister counsellor) Medium (division head) Lower (2× first secretary)
IO 6	High level (ambassador, dep. permanent representative)

Third, we conducted interviews with officials from IO administrations, also focusing on senior management staff—either at the high-level (chef de cabinet, programme heads, directors) or at the medium-level (unit heads, advisors). Again, the table below demonstrates that all IOs were covered equally. In IO 5, we did not get access to medium-level management staff, yet we were able to speak with high-level officials.

IO	Interviews with IPA officials (level, position)
IO 1	High level (Senior Programme Officer) Medium level (Programming Unit Head)
IO 2	High level (Senior Management Office) Medium level (2× Unit Head, Senior Advisor)
IO 3	High level (Division Director) Medium level (Senior Advisor)
IO 4	High level (Country Office Head, Programme Head) Medium level (Senior Manager)
IO 5	High level (2× Department Director, Senior Management Officer)
IO 6	High level (Division Director) Medium level (Senior Advisor)



### Appendix 3: Semi-structured interview questionnaire

Evaluation system*	Who and how approves the agenda of centralized evaluation system? [evaluators] Who and how approves the budget of centralized evaluation system? [evaluators] Who and how approves the staff of centralized evaluation unit? [evaluators]
Evaluation Process	How are the terms of reference formulated for centralized evaluations? [evaluators] Who conducts centralized evaluations? [evaluators] How are the reports drafted? [evaluators] Who has access to draft evaluation results? [evaluators] How do you present evaluation results to the stakeholders? [evaluators] What happens after draft evaluation reports are finalized (management response, approval of recommendations, follow-up)? [evaluators] What are the main challenges in your work? How do you deal with them? [evaluators] To what extent are you [your unit/department/mission] involved in [IO] evaluation activities (central evaluation function)? Is there any contestation about the evaluation process or results before they get officially confirmed? How satisfied are you with the process of centralized evaluations in [IO]?
Evaluation Use	In your view, what purposes does evaluation fulfil in [IO]? How would you describe the use of evaluation in [IO]? How would you describe member states/IPA interest and involvement in evaluation activities?

\*We double-checked the results of our analysis of structural evaluations factors (agenda-setting and budgetary resources) based on primary evaluation documents with the staff of evaluation units

### Appendix 4: Coded statements per IO/interviewee group

Overview of total coded statements on political evaluation use (dV2) across the three interviewee target groups in six international organizations:

	International Organizations					
	IO 1	IO 2	IO 3	IO 4	IO 5	IO 6
Member states	1	14	8	2	12	11
IPA	6	11	6	6	24	10
Evaluation Unit	11	2	6	10	10	5
Total coded statements:	18*	27	20	18*	46	26

\*In IO 1 and IO 4, member states made only few statements on political evaluation use and the total number of statements is also lower than in other organizations. In our view, this implies that the political use of evaluation is less pronounced in these organizations. Note that every target group in every IO was interviewed in a balanced way (see “Appendix 2”)

The dependent variable 1 (evaluation unit orientation) was coded based on interviews with evaluation unit staff (heads of units)

## Appendix 5: IOs' general characteristics

The three clusters include international organizations which vary in their policy fields, capacities (staff and budgetary resources), the level of operability (staff outside HQ) as well as the dependency on voluntary contributions (see table below). Given that this variation holds both within the clusters and across them, we hold that such broad IO characteristics cannot account for the observed variation in the dependent variables.

Cluster	IO	Policy field	Staff		Budget	
			Total	Decentralized (%)	Total revenue (USD)	Non-assessed (%)
1	ILO	Social	3171	70	691,926,212	44
	UNDP	Development	7013	82	5,517,025,115	100
	UNICEF	Human rights	14,474	87	6,675,758,606	100
2	IOM	Humanitarian aid	4888	94	1,862,731,880	97
	UNHCR	Humanitarian aid	10,197	93	4,338,294,301	99
	IAEA	Security	2547	3	667,839,306	38
3	WHO	Health	8153	71	2,901,381,847	83
	UNESCO	Social	2206	48	683,830,086	51
	UNEP	Environment	910	18*	741,749,419	68

Data as of 2018, based on IO websites and UN System Chief Executives Board for Coordination data (UN CEB). Policy fields coded manually based on IO websites (IOs' self-description)

\*Data as of 2010

## References

- Abbott, K. W., & Snidal, D. (1998). Why states act through formal international organizations. *The Journal of Conflict Resolution*, 42(1), 3–32.
- Anderson, J. E. (1975). *Public policy-making*. New York: Praeger.
- Azzam, T. (2010). Evaluator responsiveness to stakeholders. *American Journal of Evaluation*, 31(1), 45–65. <https://doi.org/10.1177/1098214009354917>.
- Banner, D. K. (1974). The politics of evaluation research. *OMEGA, The International Journal of Management Science*, 2(6), 763–774.
- Bauer, M. W., Knill, C., & Eckhard, S. (Eds.). (2017). *International bureaucracy. Challenges and lessons for public administration research*. London: Palgrave Macmillan.
- Beach, D., & Pedersen, R. B. (2016). Selecting appropriate cases when tracing causal mechanisms. *Sociological Methods & Research*. <https://doi.org/10.1177/0049124115622510>.
- Biermann, F., & Siebenhüner, B. (Eds.). (2009). *Managers of global change. The influence of international environmental bureaucracies*. Cambridge, MA: MIT Press.
- Bjornholt, B., & Larsen, F. (2014). The politics of performance measurement. 'Evaluation use as mediator for politics'. *Evaluation*, 20(4), 400–411. <https://doi.org/10.1177/1356389014551485>.
- Botterill, L. C., & Hindmoor, A. (2012). Turtles all the way down. Bounded rationality in an evidence-based age. *Policy Studies*, 33(5), 367–379. <https://doi.org/10.1080/01442872.2011.626315>.
- Brewer, G. D. (1974). The policy sciences emerge. To nurture and structure a discipline. *Policy Sciences*, 5(3), 239–244. <https://doi.org/10.1007/bf00144283>.
- Chelmsky, E. (1987). The politics of program evaluation. *Society*, 25, 24–32.
- Cooley, A., & Snyder, J. (Eds.). (2015). *Ranking states as a tool of global governance*. Cambridge: Cambridge University Press.

- da Conceição-Heldt, E. (2013). Do agents “run amok”? A comparison of agency slack in the EU and US trade policy in the Doha Round. *Journal of Comparative Policy Analysis: Research and Practice*, 15(1), 21–36. <https://doi.org/10.1080/13876988.2012.754152>.
- Denzin, N. K., & Lincoln, Y. S. (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks: Sage Publication.
- Dijkstra, H. (2015). Shadow bureaucracies and the unilateral control of international secretariats. Insights from UN peacekeeping. *Review of International Organizations*, 10(1), 23–41.
- Easterly, W. (2006). *The white man's burden. Why the West's efforts to aid the rest have done so much ill and so little good*. New York: Penguin Press.
- Eckhard, S., & Ege, J. (2016). International bureaucracies and their influence on policy-making. A review of empirical evidence. *Journal of European Public Policy*, 23(7), 960–978. <https://doi.org/10.1080/13501763.2016.1162837>.
- Eckhard, S., & Jankauskas, V. (2019). The politics of evaluation in international organizations. A comparative study of stakeholder influence potential. *Evaluation*, 25(1), 62–79.
- Elsig, M. (2011). Principal–agent theory and the World Trade Organization. Complex agency and ‘missing delegation’. *European Journal of International Relations*, 17(3), 495–517. <https://doi.org/10.1177/1354066109351078>.
- FAO. (2016). *Evaluation of FAO's evaluation function. PC 120/5*. Rome: Food and Agriculture Organization of the United Nations.
- Fforde, A. (2019). Yes, but what about the authority of policy analysts? A commentary and discussion of Perl et al., ‘Policy-making and truthiness: Can existing models cope with politicized evidence and willful ignorance in a post-fact world?’. *Policy Sciences*, 52(1), 153–169. <https://doi.org/10.1007/s11077-018-9344-2>.
- George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press (BCSIA studies in international security).
- Goffman, E. (1959). *The presentation of the self in everyday life*. New York: Doubleday.
- Graham, E. R. (2014). International organizations as collective agents. Fragmentation and the limits of principal control at the World Health Organization. *European Journal of International Relations*, 20(2), 366–390. <https://doi.org/10.1177/1354066113476116>.
- Hanrieder, T. (2015). *International organization in. Time fragmentation and reform*. Oxford: Oxford University Press.
- Hawkins, D. G., Lake, D. A., Nielson, D. L., & Tierney, M. J. (Eds.). (2006). *Delegation and agency in international organizations*. Cambridge: Cambridge University Press.
- Hinterleitner, M., Sager, F., & Thomann, E. V. A. (2016). The politics of external approval. Explaining the IMF's evaluation of austerity programmes. *European Journal of Political Research*, 55(3), 549–567. <https://doi.org/10.1111/1475-6765.12142>.
- Højlund, S. (2014). Evaluation use in the organizational context—Changing focus to improve theory. *Evaluation*, 20(1), 26–43. <https://doi.org/10.1177/1356389013516053>.
- Højlund, S. (2015). Evaluation in the European Commission. For accountability or learning? *European Journal of Risk Regulation*, 6(1), 35–46.
- Howlett, M., Ramesh, M., & Perl, A. (2009). *Studying public policy. Policy cycles and policy subsystems*. Oxford: Oxford University Press.
- IEO. (2018). *Annual report on evaluation. Independent evaluation office*. New York: United Nations Development Programme.
- ILO. (2019). *Annual evaluation report 2018–2019*. Geneva: International Labour Organization.
- JIU. (2014). Analysis of the Evaluation Function in the United Nations System. United Nations Joint Inspection Unit. JIU/REP/2014/6. Geneva.
- Johnson, T., & Urpelainen, J. (2014). International bureaucrats and the formation of intergovernmental organizations. Institutional design discretion sweetens the pot. *International Organization*, 68(1), 177–209. <https://doi.org/10.1017/s0020818313000349>.
- Kassim, H., & Menon, A. (2003). The principal-agent approach and the study of the European Union. Promise unfulfilled? *Journal of European Public Policy*, 10(1), 121–139. <https://doi.org/10.1080/1350176032000046976>.
- Knill, C., Bayerlein, L., Enkler, J., & Grohs, S. (2019). Bureaucratic influence and administrative styles in international organizations. *Review of International Organizations*, 14, 83–106. <https://doi.org/10.1007/s11558-018-9303-x>.
- Koremenos, B., Lipson, C., & Snidal, D. (2001). The rational design of international institutions. *International Organization*, 55(4), 761–799. <https://doi.org/10.1162/002081801317193592>.
- Lall, R. (2017). Beyond institutional design. Explaining the performance of international organizations. *International Organization*, 71(2), 245–280. <https://doi.org/10.1017/s0020818317000066>.

- Lasswell, H. D. (1956). *The decision process. Seven categories of functional analysis*. College Park: Bureau of Governmental Research: University of Maryland Press.
- Leeuw, F. L., & Furubo, J.-E. (2008). Evaluation systems. What are they and why study them? *Evaluation*, 14(2), 157–169. <https://doi.org/10.1177/1356389007087537>.
- Lijphart, A. (1971). Comparative politics and the comparative method. *American Political Science Review*, 65, 682–693.
- Lyne, M. M., Nielson, D. L., & Tierney, M. J. (2006). Who delegates? Alternative models of principals in development aid. In D. G. Hawkins, D. A. Lake, D. L. Nielson, & M. J. Tierney (Eds.), *Delegation and agency in international organizations* (pp. 41–76). Cambridge: Cambridge University Press.
- Mayntz, R. (2004). Mechanisms in the analysis of social macro-phenomena. *Philosophy of the Social Sciences*, 34(2), 237–259. <https://doi.org/10.1177/0048393103262552>.
- Mele, V., Anderfuhren-Biget, S., & Varone, F. (2016). Conflicts of interest in International Organizations. Evidence from two United Nations humanitarian agencies. *Public Administration*, 94(2), 490–508. <https://doi.org/10.1111/padm.12228>.
- Merry, S. E. (2011). Measuring the world. *Current Anthropology*, 52(S3), S83–S95. <https://doi.org/10.1086/657241>.
- Morris, M., & Clark, B. (2013). You want me to do what? Evaluators and the pressure to misrepresent findings. *American Journal of Evaluation*, 34(1), 57–70. <https://doi.org/10.1177/1098214012457237>.
- Nielson, D. L., & Tierney, M. J. (2003). Delegation to international organizations. Agency theory and World Bank environmental reform. *International Organization*, 57(2), 241–276. <https://doi.org/10.1017/s0020818303572010>.
- O'Brien, T., Payne, S., Nolan, M., & Ingleton, C. (2010). Unpacking the politics of evaluation. A dramaturgical analysis. *Evaluation*, 16(4), 431–444. <https://doi.org/10.1177/1356389010380002>.
- Patz, R., & Goetz, K. H. (2019). *Managing money and discord in the UN. Budgeting and bureaucracy*. Oxford: Oxford University Press.
- Perl, A., Howlett, M., & Ramesh, M. (2018). Policy-making and truthiness. Can existing policy models cope with politicized evidence and willful ignorance in a “post-fact” world? *Policy Sciences*, 51(4), 581–600. <https://doi.org/10.1007/s11077-018-9334-4>.
- Pleger, L., Sager, F., Morris, M., Meyer, W., & Stockmann, R. (2017). Are some countries more prone to pressure evaluators than others? Comparing findings from the United States, United Kingdom, Germany, and Switzerland. *American Journal of Evaluation*, 38(3), 315–328. <https://doi.org/10.1177/1098214016662907>.
- Porter, T. M. (1995). *Trust in numbers. The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.
- Raimondo, E. (2018). The power and dysfunctions of evaluation systems in international organizations. *Evaluation*, 24(1), 26–41. <https://doi.org/10.1177/1356389017749068>.
- Rittberger, V., Zangl, B., Kruck, A., & Dijkstra, H. (2019). *International organization* (3rd ed.). London: Red Globe Press.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (Eds.). (2004). *Evaluation. A systematic approach* (7th ed.). Thousand Oaks: Sage Publication.
- Schoenefeld, J., & Jordan, A. (2017). Governing policy evaluation? Towards a new typology. *Evaluation*, 23(3), 274–293. <https://doi.org/10.1177/1356389017715366>.
- Schoenefeld, J., & Jordan, A. (2019). Environmental policy evaluation in the EU. Between learning, accountability, and political opportunities? *Environmental Politics*, 28(2), 365–384. <https://doi.org/10.1080/09644016.2019.1549782>.
- Stockmann, R., Meyer, W., & Schenke, H. (2011). Independence of evaluations. *Zeitschrift für Evaluation*, 10(1), 39–67.
- Streck, C. (2001). The global environment facility. A role model for international governance? *Global Environmental Politics*, 1(2), 71–94.
- Taylor, D., & Balloch, S. (Eds.). (2005). *The politics of evaluation. Participation and policy implementation*. Bristol: The Policy Press.
- The LSE GV314 Group. (2014). Evaluation under contract Government pressure and the production of policy research. *Public Administration*, 92(1), 224–239. <https://doi.org/10.1111/padm.12055>.
- UNDP. (2016). *The UNDP evaluation policy. DP/2016/23*. New York: United Nations Development Programme.
- UNEG. (2016). *Norms and standards for evaluation*. New York: United Nations Evaluation Group.
- UNHCR. (2016). *UNHCR policy on evaluation. UNHCR/HCP/2016/2*. Geneva: United Nations High Commissioner for Refugees.
- Urpelainen, J. (2012). Unilateral influence on international bureaucrats. An international delegation problem. *Journal of Conflict Resolution*, 56(4), 704–735. <https://doi.org/10.1177/0022002711431423>.

- van Voorst, S., & Mastenbroek, E. (2019). Evaluations as a decent knowledge base? Describing and explaining the quality of the European Commission's ex-post legislative evaluations. *Policy Sciences*, 52(4), 625–644. <https://doi.org/10.1007/s11077-019-09358-y>.
- Vaubel, R., Dreher, A., & Soylu, U. (2007). Staff growth in international organizations. A principal-agent problem? An empirical analysis. *Public Choice*, 133(3), 275–295. <https://doi.org/10.1007/s11127-007-9188-3>.
- Voeten, E. (2008). The impartiality of international judges. Evidence from the European Court of Human Rights. *American Political Science Review*, 102(4), 417–433.
- Weaver, C. (2010). The politics of performance evaluation. Independent evaluation at the International Monetary Fund. *Review of International Organizations*, 5(3), 365–385.
- Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19(1), 21–33.
- Weiss, T. G., & Jordan, R. S. (1976). Bureaucratic politics and the world food conference. The international policy process. *World Politics*, 28(3), 422–439. <https://doi.org/10.2307/2009978>.
- Weiss, C. H., Murphy-Graham, E., & Birkeland, S. (2005). An alternate route to policy influence. How evaluations affect D.A.R.E. *American Journal of Evaluation*, 26(1), 12–30. <https://doi.org/10.1177/1098214004273337>.
- Wergin, J. F. (1976). The evaluation of organizational policy making. A political model. *Review of Educational Research*, 46(1), 75–115. <https://doi.org/10.3102/00346543046001075>.
- WHO. (2018). *Evaluation. Evaluation Policy. EB143(9)*. Geneva: World Health Organization.
- Wildavsky, A. (1972). The self-evaluating organization. *Public Administration Review*, 32(5), 509. <https://doi.org/10.2307/975158>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.