Erschienen in: Evaluation ; 25 (2019), 1. - S. 62-79
https://dx.doi.org/10.1177/1356389018803967

Article

Evaluation
2019, Vol. 25(1) 62–79

DOI: 10.1177/1356389018803967
journals.sagepub.com/home/evi
SAGE

The politics of evaluation in international organizations: A comparative study of stakeholder influence potential

Steffen Eckhard
University of Konstanz, Germany

Vytautas Jankauskas
Ludwig-Maximilians-University Munich, Germany
University of Konstanz, Germany

Abstract
While the political nature of evaluation is widely recognized, few attempts exist to conceptualize and compare these politics. This article develops the concept of evaluation stakeholder influence potential, which builds on four political resources for influence (agenda-setting powers, staff and budgetary resources, access to evaluation results, and access to evaluators). These resources are measured for both member states and international public administrations in 24 United Nations organizations. We find that the administration—and not member states—have the largest influence potential in almost two-thirds of the international organizations. Our findings allow classifying them into three groups for which we expect differences in political contestation about evaluation use: two extreme-case groups (either member state or administrative dominance) and a group of contested middle cases. This finding of bureaucratic dominance reinforces literature on bureaucrats as powerful evaluation stakeholders in domestic settings and speaks to nascent research on bureaucratic influence in international organizations.

Keywords
bureaucracy, evaluation politics, influence potential, international organizations, stakeholder influence

Introduction

This article introduces a new concept to measure stakeholders' potential to influence evaluation results and applies it to a set of 24 United Nations (UN) organizations. Doing so, this

Corresponding author:
Steffen Eckhard, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany.
Email: steffen.eckhard@uni-konstanz.de

article taps into a long-standing debate on the politics of evaluation. This literature recognizes that despite its ambitious intention to be a scientific and data-based inquiry, evaluation is an inherently political activity (Azzam, 2010; Banner, 1974; Bjornholt and Larsen, 2014; Brandon and Fukunaga, 2014; Chelimsky, 1987; Cohen, 1972; House, 1974; Malen et al., 1988; Morris and Clark, 2013; O'Brien et al., 2010; Orr, 2010; Picciotto, 2016; Raimondo, 2018; Taylor and Balloch, 2005; The LSE GV314 Group, 2014; Weiss, 1993; Weiss and Jordan, 1976).

Despite the broad scholarly consensus that interest contestation makes evaluation political, the existing literature does not offer a robust framework to compare how stakeholders can influence evaluation processes and how this varies between organizations. This article addresses two important research gaps in particular. First, there is still a lack of conceptual approaches to the study of stakeholder influence on evaluation processes. Most studies are based on single-case observations and tend to merely recognize stakeholders' involvement without generalizing it conceptually (e.g. Weiss, 1999; Wildavsky, 1972). For example, studies highlight that 'being pressured to misrepresent findings is a common occurrence in evaluation' (Morris and Clark, 2013: 66) and that '[s]takeholders can dramatically affect how an evaluation is designed and implemented' (Azzam, 2010: 45). It remains unclear what empowers stakeholders to affect the results of evaluation and how can we measure it. Second, the overwhelming majority of studies on the politics of evaluation primarily focus on the national level.[1] By contrast, researchers increasingly call for more research on evaluation at the international level (Hojlund, 2014: 428) and in international organizations (IOs) in particular. As Raimondo (2018: 29) highlights, 'to date there is no single body of literature that can satisfactorily explain the role and performance of evaluation systems in International Organizations.'

Lack of attention to evaluation in IOs is unfortunate, considering the increasing institutionalization of evaluation and its overall significance for IO policymaking (Chelimsky and Shadish, 1997; Joint Inspection Unit (JIU), 2014a: 12).[2] Studies highlight that IO policymaking is contested among member states and that IO bureaucracies—the international public administrations (IPAs)—exert autonomous influence on policymaking (for an overview, see Eckhard and Ege, 2016). In this politicized setup, it is likely that political and bureaucratic evaluation stakeholders with deviating conceptions of organizational goals and what constitutes success or failure seek to influence evaluation. Combining IO and evaluation literature, this article thus scrutinizes the UN system's own definition of 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' (United Nations Evaluation Group (UNEG), 2016: 10, emphasis added).

Against this backdrop, the article first develops the concept of evaluation stakeholder influence potential, which can be applied to evaluation systems both in domestic and international public organizations. Drawing on a resource-based view of political power (Clark, 1968; Lowery, 2013; Mills, 1956; Simon, 1953), the proposed taxonomy (formal vs informal processes; direct vs indirect access) maps four different categories of political resources for influence: agenda-setting powers, staff and budgetary resources, access to (draft) evaluation results, and access to evaluators themselves. These resources can be attributed to different stakeholders and thus demonstrate which of them has the largest potential to influence evaluation results. We operationalize the concept for the context of UN system IOs and draw on empirical data provided by the UN JIU (2014a) to advance a comparative measurement of the influence potential of key IO stakeholders—both member states and IPAs—across 24 UN organizations. While systemic peculiarities make comparison of evaluation systems at the national level

unusually difficult (but see Jacob et al., 2015; Pleger et al., 2016), IOs—as relatively equal political systems consisting of a member state's collective and a bureaucracy—constitute a comparable class of cases.

This way, the article contributes to the literature on evaluation and IOs in three ways. First and conceptually, we propose a resource-based heuristic for the comparative study of stakeholder influence potential. This provides for a better understanding of how organizational evaluation systems vary and by whom evaluation processes can potentially be influenced (Dahler-Larsen, 2012; Hojlund, 2014; Leeuw and Furubo, 2008; Raimondo, 2018; Rutkowski and Sparks, 2014). Given its demonstrated utility in the context of IOs, future research might find it useful to also apply the stakeholder influence potential concept in a domestic context.

Second and empirically, this article is the first to demonstrate the variation of stakeholder influence potential across 24 UN organizations. This allows for classifying IO evaluation systems into three groups for which we expect different evaluation politics when stakeholders realize their influence potential. Since the patterns of stakeholder influence arguably affect evaluation use (see Weiss, 1993), our study provides plausible points of departure for future empirical studies on the use of evaluation in concrete IOs.

Third and theoretically, the analysis yields that IPAs—and not member states—have the largest potential to influence evaluation results in almost two-thirds of the IOs. This finding speaks to recent studies on evaluation in IOs, which put forth the claim that evaluation is 'positioned within a global governance strategy that seeks greater influence for International Organizations' (Raimondo, 2018: 34; see also Rutkowski and Sparks, 2014). More precisely, as in the domestic context (Bjornholt and Larsen, 2014; Leeuw and Furubo, 2008; Weiss, 1993; Wildavsky, 1972), bureaucrats within IOs have the potential to be powerful evaluation stakeholders who could use evaluation as a tool to steer principals and what is politically achievable. This finding on bureaucratic dominance over evaluation also reinforces a nascent research on IPAs' influence on IO policymaking (Bauer et al., 2017; Biermann and Siebenhüner, 2009; Eckhard and Ege, 2016; Goetz and Patz, 2017; Hooghe and Marks, 2015; 2016; Knill et al., 2016, 2018; Trondal et al., 2010).

The politics of evaluation in international organizations

The political nature of evaluation was already mentioned in the early 1970s, when one of the first articles on this topic was written by Weiss (1970; see also Banner, 1974; Cohen, 1972; Weiss and Jordan, 1976). Some saw 'dirty politics' as a threat and 'intrusion into scientific evaluation' (Davidson, 2005; House, 2008; Scriven, 1991: 268), while others treated it as an inevitable reality that must be accepted in order to increase evaluation use[3] (Patton, 1997; Weiss, 1993).

But what makes evaluation political? In general, evaluation is conducted when its stakeholders require specific information about the evaluation object. Evaluation results are thus thought to objectively inform stakeholders and accordingly alter their behavior (Bjornholt and Larsen, 2014: 408; Heinrich, 2002: 713). Yet for this very reason, the stakeholders themselves may seek to favorably impact evaluation results. For example, evaluation reports may be used for resource redistribution, which in turn affects power relationships within the organization, making evaluation an activity with vested interests (Wergin, 1976: 76). In other words, evaluation is 'an interactive process' (Lee, 2006: 170) in which involved actors 'articulate their interests from different positions of power, influence, and authority' (Datta, 2011: 274). Thus,

in order to build a comprehensive understanding of evaluation, one should not limit its meaning to a mere 'audit logic' but rather see it as a system (Taylor and Balloch, 2005: 109; see Leeuw and Furubo, 2008). Although scientific methods and technical issues are important, both evaluations and their objects 'do not take place in a vacuum' (Raimondo, 2018: 26); they exist in a political environment (see Cronbach et al., 1981; Rossi et al., 2004: 48; Schoenefeld and Jordan, 2017: 289; Weiss, 1972). Actors involved in evaluation—its clients, users, and stakeholders—are usually at the same time embedded in the process of decision-making, and their evaluation-related interests emerge 'as a manifestation of the political nature of evaluation' (Brandon and Fukunaga, 2014: 27).

Assuming that stakeholders seek to influence evaluation results for their own preference, one has to clarify who these stakeholders are and what evaluation-related interests they have. In the national context, scholars have identified a number of different stakeholder groups. Picciotto (2016), for example, differentiates between the commissioning entity, the evaluator, decision makers, and beneficiaries. Rossi et al. (2004) add evaluation and research community, contextual stakeholders, program competitors, and program, as well as evaluation sponsors (pp. 48–49). However, the potential to influence evaluation varies across these actors. As evaluation research indicates, stakeholders with the largest influence are 'the political and administrative elite,' such as national authorities and civil servants (Bjornholt and Larsen, 2014: 409; Weiss, 1993: 93), in other words, actors who are directly involved in the decision-making process of the evaluated program (see Azzam, 2010: 47; Chelimsky, 1987; Roker, 2005: 123). Case studies in different countries have clearly emphasized that those who ordered the evaluation were 'the primary influencing stakeholder groups' (Pleger et al., 2016: 315). While *decision makers* were observed to be extremely powerful 'due to their direct control over logistical factors such as evaluation funding, timeline, and staff support' (Azzam, 2010: 47), *bureaucracy* was also identified as an influential actor due to its embeddedness in the evaluation system (Leeuw and Furubo, 2008; The LSE GV314 Group, 2014).

Turning to the international level, IOs appear to serve as extreme cases to study political and administrative stakeholder influence. First, key *decision makers* of organizations—the member states—constitute much more complex principals than, for instance, national parliaments at the domestic level. The IO membership often includes almost 200 member countries and is characterized by high interest heterogeneity (Eckhard and Dijkstra, 2017; Hawkins et al., 2006). In policymaking, powerful states often realize their interests through informal lobby-like mechanisms, whereas less powerful states depend on formal instruments, such as evaluation, to counter such lobbying (Stone, 2011). Second, the IO bureaucracy, or IPA, is less constrained by such 'stabilizers' as political party competition or national public administration law and is more likely to develop its own interests (Bauer and Ege, 2017). Previous studies found IPAs to be highly entrepreneurial and innovative in devising new administrative tools and strategies to influence policymaking (Bauer et al., 2017; Biermann and Siebenhüner, 2009; Eckhard and Ege, 2016; Goetz and Patz, 2017; Hooghe and Marks, 2015; Jörgens et al., 2016; Knill et al., 2016, 2018; Trondal et al., 2010). This means that evaluation-related interests multiply, making IO stakeholder influence a crucial topic to investigate with potential implications even beyond IOs.

A number of motives for stakeholder evaluation influence can be derived from evaluation and IO literature. First, member states or IPAs may seek to get favorable evidence for a particular policy, either to justify its existence or to yield support for (no-)action (see Weiss, 1999). As an example, the study by Morris and Clark (2013) concluded that '[m]

isrepresentation pressure usually focused on making evaluation results look more positive or less negative than the evaluator thought was warranted' (p. 57). Stakeholders may also want to influence evaluation results to use it as a bargaining chip for resource distribution (Wergin, 1976), for blame-shifting dynamics, or to avoid own policy commitments (Bjornholt and Larsen, 2014: 403).

Second, and more generally, member states may want to hold IPAs accountable and thus shape evaluation results in a way that allows them to put bureaucracies under hierarchical control. It is one of the premises of principal–agent theory that bureaucratic agents attempt to elude the control of their principals because they want to do less or differently than asked (Elsig, 2011; Vaubel et al., 2007). Therefore, member states are functionally compelled to apply control mechanisms, evaluation being one of them. In contrast, IO literature increasingly acknowledges that IPAs may seek to deliberately escape member states' control to increase autonomy. Thus, IPAs may influence evaluation results to hide information (adverse selection or ex ante opportunism) or action (moral hazard or ex post opportunism) (see da Conceição-Heldt, 2013: 24).

It is reasonable to conclude that significant contestation about evaluation processes should exist in IOs. On one hand, evaluation literature yields that there is hardly any organizational context without stakeholder interests about the results of evaluation. On the other hand, while domestic political stakeholders, such as the parliament versus government, also compete, IPA influence and member state competition at the international level render IO policymaking particularly complex. But it is still an open question whose interest evaluation serves and to what end. As a first step toward answering these questions, we propose a conceptual framework for the comparative study of evaluation stakeholders' influence potentials. Observing variation in stakeholder influence speaks to the literature on IO policymaking. But the concept aims to be broad enough also to apply to other institutional settings beyond IOs alone.

The concept of evaluation stakeholder influence potential

While some studies analyze how evaluators react to stakeholders' pressure (e.g. Azzam, 2010), we focus on the potential influence of stakeholders themselves. In other words, which stakeholder group has the largest potential to influence evaluation results and why? The term 'influence' is used as synonymous with 'having an effect' (Biermann and Siebenhüner, 2009: 40; Eckhard and Ege, 2016: 964; Liese and Weinlich, 2006: 504). Thus, it refers to a causal reasoning of one actor changing the behavior of the other (Cox and Jacobson, 1973: 3). It is about evaluation stakeholders having an effect on evaluators' behavior with regard to evaluation results (i.e. both its findings and recommendations). Importantly, influence *potential* is not so much about causal claims that A has an effect on B, but rather about the opportunity to exercise such influence (i.e. to have an effect). Hence, we focus on the potential of influence as a matter of capability in contrast to the influence itself as a matter of a causal act.

Our understanding of influence potential draws on a resource-based approach to the study of political power (Clark, 1968: 46; Lowery, 2013: 3; Mills, 1956; Simon, 1953). According to this perspective, one has to differentiate 'between the *exercise* of power—when A induces, forces or compels B to do something that B would not otherwise do—and the bases of power or the resources that allow for its exercise' (Lowery, 2013: 3). As Cox and Jacobson (1973) put it, '[p]ower means capability; it is the aggregate of political resources that are available to an actor … Power may be converted into influence, but it is not necessarily so' (p. 4). Therefore,

Table 1. Mapping of stakeholders' political resources to potentially influence evaluation results.

		Participation in evaluation process	
		Direct	Indirect
Social interaction	Formal	*Agenda-setting powers*	*Resources (staff and budget)*
	Informal	*Access to evaluation results*	*Access to evaluators*

stakeholders with most political resources—or bases of influence—are to be considered as the most powerful stakeholders, that is, the ones with the largest potential to influence evaluation results.

In relation to this, what are the typical stages in the evaluation process and what political resources can be used to influence them? First, conducting evaluation begins with several formal steps to get evaluators working. This is what we refer to as an evaluation initiation phase. The program of work has to be decided (agenda setting) and both personnel (staff) and material (budget) resources have to be allocated to the evaluation unit. The second phase is about the narrowing down of evaluation questions and scope (terms of reference), evaluation researchers must be found, methods and data sources agreed (scoping study), and the research must be conducted. Ultimately, the drafting and finalizing phase refers to the subsequent consultations (often in the informal), quality control, and the formulation of recommendations and the production of the final report.[4]

In this regard, several studies have dealt with the question of how stakeholders are involved in the process of evaluation and how they can seek to realize political interests (Azzam, 2010; Morris and Clark, 2013; Morris and Cohn, 1993). The described stages emerge as opportunities for stakeholders to potentially influence the final output, although the level of contestation as well as the type of political resources used at different stages differs. Building on the existing findings on stakeholder influence, the key political resources that stakeholders may have can be (1) based on either formal or informal social interactions and they can be (2) used through either direct or indirect participation in the evaluation process. As summarized in Table 1 above, the combination of formal or informal interaction and the direct or indirect participation allows distinguishing four measurable political resources for stakeholders to potentially influence evaluation results.

First, all processes in organizations (be they national or international) can be perceived as social and consisting of both formal and informal elements (March and Olsen, 1989, 1998). Formal political resources refer to institutionalized rule structures (Scott, 2007). One of the widely recognized formal resources are stakeholders' *agenda-setting powers*, which can be used through a direct participation in the phase of evaluation initiation. Usually, evaluators are not free to decide upon the evaluation work plan. Instead, 'they have to work with the evaluation sponsor, program management, and other stakeholders to develop this essential background' (Rossi et al., 2004: 41). The very ability to decide what is (or is not) to be evaluated gives stakeholders the power of shaping the output of evaluation. As Bjornholt and Larsen (2014: 404) imply, evaluation's program of work reflects stakeholders' interests and predetermines what kind of answers can be delivered.

Furthermore, stakeholders can use their formal leverage over evaluation *resources*, such as staff and budgets, in order to indirectly (i.e. acting outside the actual process of evaluation) influence the results of evaluation. A simulation study by Azzam (2010) 'revealed 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' (p. 45). Similarly, Rossi et al. (2004) noticed that funding is 'a critical resource around which the evaluation must be planned' (p. 46). Stockmann et al. (2011) also highlighted the various opportunities for evaluation's funders to manipulate its process and output. Regarding evaluation staff, Morris and Clark (2013) demonstrated that most often, 'the pressure to misrepresent came from the individual who hired the respondent or assigned him or her to conduct the evaluation' (p. 61). Traditionally, principal–agent approaches treated both staff and money as valuable principals' (stakeholders') resources for sanctioning purposes in order to control agents (evaluators) (Hawkins et al., 2006). Just as managers in a firm seek to have good relations with their bosses, evaluators as agents usually respond to their principals' preferences (O'Brien et al., 2010: 432). By setting the budget or selecting favorable evaluators, stakeholders can try to shape their behavior and thus affect an evaluation's output even if not directly participating in further evaluation process.

Second, stakeholders may also have informal political resources to affect evaluation results (March and Olsen, 1989, 1998). Scholars tend to notice that in organizational decision-making, especially at the international level, 'informality is omnipresent' (Lipson, 1991: 495). Informal resources imply that stakeholders affect evaluation not through formal rule structures, but rather through non-institutionalized social interaction such as persuasion, positive or negative inducement, or deception (Banfield, 1961: 4; Bohne, 2010). As Stone (2011) has argued, '*informal influence* consists of participation in decision making and special access to information' (p. 33; emphasis in original). In influence literature, access is indeed recognized as a significant but at the same time 'scarce resource' (see Lowery, 2013: 6).

Against this backdrop, having *access to evaluation results* (i.e. through direct participation in the evaluation process) before they get officially confirmed or published gives one the ability to use lobby-like strategies to translate own preferences into the final output during the drafting and finalizing stage. For instance, stakeholders who have access to preliminary evaluation findings may withhold important information required for the further inquiry or apply 'more subtle' forms of pressure by providing suggestive content alterations (see Pleger et al., 2016: 323). The simulation study by Azzam (2010) showed that a mere meeting with the stakeholders during the evaluation process to hear their feedback on evaluation design led evaluators to modify the procedures afterward. This can be perceived as 'windows of opportunity' for stakeholders to move their preferred items into the evaluation's output, to impact the interpretation of already gathered data, or to shape the proposed recommendations (Hardt, 2013: 338; Howlett et al., 2009: 135).

Related but nonetheless distinct is the ability of stakeholders to *access evaluators* themselves, even being outside the actual evaluation process (i.e. through indirect participation). It is distinctive because stakeholders other than those who are directly involved in consulting draft results can exert informal influence (see Urpelainen, 2012). For instance, political actors may not have formal access to the actual draft results but may instead seek to privately meet evaluators or evaluation managers. They may, for instance, apply lobby-like strategies to discuss the objectives of evaluations, elements of the terms of reference, or methodological choices. Stakeholders' ability to gain such informal access varies depending on the organizational structure and their own capacities (e.g. Graham, 2014).

As Figure 1 summarizes, the four political resources constitute an opportunity structure to influence evaluation processes directly or indirectly. While the power over agenda setting and

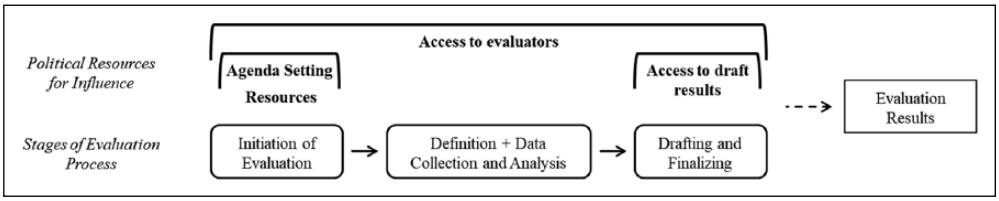

Figure 1. Political influence resources during evaluation process.

evaluation resources might be exercised mainly during the initiation phase, access to (draft) results might be used to influence evaluation during the drafting and finalizing stage. Finally, the ability to informally access evaluators is salient throughout the whole process both when setting the overall evaluation work plan and when conducting the evaluation research.

Evaluation processes are undoubtedly complex and so are IOs' institutional designs. In this sense, the concept of stakeholder influence potential is limited in its ability to capture the full diversity of individual evaluation processes. Instead, it allows observing general patterns of influence potential in evaluation systems of different organizations. While being aware of its limitations, the resource-based perspective offers a theory-grounded trade-off between abstraction (comparability) and precision. Whether political resources will be used for actual (causal) influence is beyond the scope of this study. Yet, even though informal access or formal capabilities of agenda setting and evaluation resources do not necessary imply influence as an act, they are crucial conditions for someone seeking to realize his or her own political interests (see Lowery, 2013: 4).

Research design and operationalization

Since this article aims to provide a comparative perspective on stakeholder influence potential across a number of IOs, the operationalization of the concept requires pragmatic, yet logically flawless, decisions to allow comparison, considering the data availability. In the following, we first describe every step of our empirical study and then discuss both the benefits and the limitations of this approach.

First, we include all IOs within the UN system that have central evaluation units[5] in our analysis (i.e. 24 IOs, see Figure 2).[6]

Second, we differentiate between two key stakeholder groups in IOs: member states and IPAs. Thus, it is the influence potential of either member states or IPAs at the level of evaluation systems that constitutes the unit of empirical analysis. The term 'member states' may have different meanings in different IOs. On the actor level, we primarily refer to member states' delegated representatives who constitute IO governing bodies (executive boards, board of governors, councils, etc.). In the case of IPAs, stakeholders are the secretary general (as the head of administration) and the heads of units/programs and their subordinates. We do not claim that other actors, such as IOs' governance targets (e.g. domestic publics), cannot seek to translate their interests or are completely neutral when it comes to evaluation processes. However, given that member states and IPAs are directly involved in IO decision-making and are therefore the commissioning entities, program sponsors, evaluation funders, and primary users at the same time, they are the usual suspects to look at whether one wants to identify political interest contestation.[7]

Table 2. Operationalization of key IO stakeholders' influence potential (based on JIU (2014a) data).

Political resource	Observable implications	Output (who has the resource)
Agenda-setting powers	Program of work decided up and issued directly to member states (and getting approval from member states)	Yes: Member states No: IPA
Resources (budget; staff)	Budgetary requests (by the evaluation unit) made directly to member states with no management interference	Yes: Member states No: IPA
	Appointment (of evaluation unit's head) by executive head with input from governing bodies	Yes: Member states No: IPA
Access to evaluation results	Evaluation report issued directly to member states without management involvement	Yes: Member States No: IPA
Access to evaluators	Evaluation unit is a stand-alone function in the IO	Yes: Member states No: IPA

IO: international organization; IPA: international public administration.

Third, we operationalize the four categories of stakeholders' political resources by drawing on data provided by the UN JIU, an independent external oversight body of the UN system mandated to conduct system-wide evaluations, inspections, and investigations. Between 2013 and 2014, two inspectors of this unit, Sukai Prom-Jackson and George A. Bartsiotas, gathered evidence on the evaluation function (evaluation methods and maturity) in all UN agencies, using expert interviews and a survey. The results were published in a 2014 report (JIU, 2014a). We contacted the authors of the report and on 30 November 2017 conducted a telephone interview with one of them to clarify their methodology and definitions. Drawing on data provided in the report and its annex, we then allocated the four political resource categories either to member states or IPAs, coded as (independent) binary variables, based on the wording and data provided by the report (see Table 2).

To begin with agenda-setting capabilities as a formal political resource, we included a category called 'agenda,' which is a binary variable answering whether evaluation's 'program of work [is] decided up and issued directly to member States (and getting approval from member States)' (JIU, 2014a: 34). If the answer to the question was positive, we perceived member states to have the agenda-setting power; otherwise, it was the IPA that had the respective political resource.

Second, in order to identify who has the control over the evaluation-related resources, we distinguished between an evaluation's funding (budgetary resources) and an evaluation's staff (personnel resources). Regarding the former, we included a category called 'budget,' which refers to the question of whether '[b]udgetary requests [by the evaluation unit] [are] made directly to member States with no management interference' (JIU, 2014a: 34). With regard to an evaluation's staff, a category labeled 'staff' depicts whether the appointment of an evaluation's unit head is made by an executive head with (or with no) input from governing bodies (i.e. member states) (JIU, 2014a: 37).

Turning to informal political resources, we included a category named 'access to results.' It refers to the question of whether an 'evaluation report [is] issued directly to member States without management involvement' (JIU, 2014a: 34). If this was the case, member states have an opportunity to discuss and potentially alter evaluation results before they get officially

confirmed. Otherwise, it is rather the IPA that is better able to translate its own interests before the report goes to the member states.

Finally, the last political resource relates to the access to evaluators rather than the draft results. Graham (2014) developed a general argument regarding IO fragmentation and member states' ability to oversee and access individual administrative units. This work implies that the more 'hidden' the evaluation function in an organization is, the harder it gets for member states to access the evaluators. To be sure, as we focus on centralized evaluation function, we primarily mean access to IPA officials working in evaluation units. They may conduct evaluations by themselves or act as evaluation managers overseeing external experts. Regardless of which 'business model' is chosen, member states may more easily reach those who are responsible for the IO evaluation system if they operate in a structurally distinct unit rather than in offices co-located with substantive units, oversight, audit, or management.[8] This is because member states often lack capacities in terms of time, staff, and information vis-à-vis IPAs. Based on the JIU data (JIU, 2014a: v, 15), we hence included a variable 'access to evaluators,' which captures member states' ability to informally interact with evaluators because of the organizational allocation of evaluation offices. In case of a stand-alone evaluation unit, the political resource of informal access is allocated to the member states. In case of co-located evaluation units, it is the IPA that benefits from its structural embeddedness and is assumed to have better access to influence evaluation officers.

The method, operationalization, and our main data source have both benefits and limitations. First, studying stakeholder influence is a challenging task and ideally requires qualitative in-depth analysis. Yet, a comparative study provides a more general understanding of a phenomenon of interest. We explicitly see our study as a first step that allows meaningful case selection for further work. Second, the evaluation processes in UN system IOs vary regarding their individual designs and complexity. For instance, the second phase of evaluation process may include additional steps that allow for influence and that our measurement neglects.[9] However, we claim that the resource-based understanding of political power underlying the identification of comparable structural proxies captures the most important variables given the phenomenon in question. The four political resources exist in all IOs, thus enabling comparison of stakeholder influence potential across evaluation systems despite diverse particulars at the level of individual evaluation processes. Future contributions can build on this and shed more light on microdynamics through expanding the scope of operationalization in a qualitative small-*n* study. Third, in drawing on a JIU report that is now a few years old, our study presents the state of the UN evaluation system as of 2014. This is unproblematic for our primary purpose of illustrating variance, but it must be kept in mind for future work, in particular, case selection. It also sets an interesting basis for future studies that could update our findings and analyze how and why stakeholder influence potential shifts over time. In addition, given that the JIU evaluation report is so far the only source for such a comparative large-*n* study, future efforts to triangulate the data by conducting interviews with the respective evaluation offices might be beneficial.

Measurement and discussion of findings

Based on the observable implications above and on the data of the JIU report (2014a), Figure 2 summarizes for 24 IOs which stakeholder controls which of the four categories of political resources for evaluation stakeholder influence. Each resource category is treated equally,

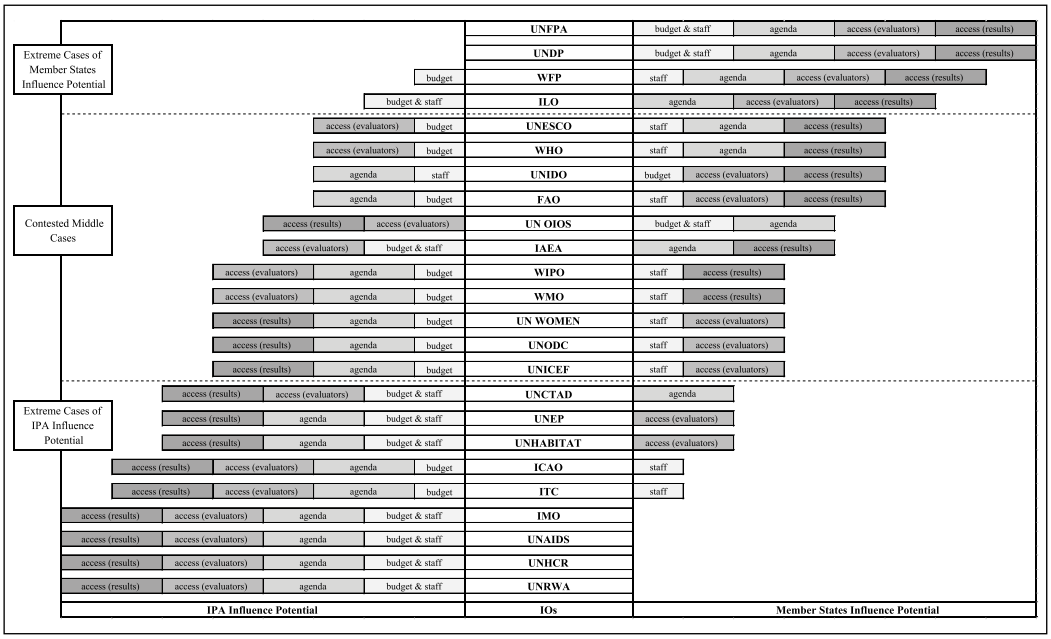

Figure 2. Varying influence potential between key evaluation stakeholders in 24 IOs (as of 2014).

which we expressed by means of the size of the boxes in Figure 2. The category of evaluation resources consists of both control over evaluation staff and the evaluation budget. In most cases, these resources are not controlled by the same actor; thus, in case of a divide, we reduced the size of the boxes by 50 percent. The resulting variation both across IOs and the two stakeholder groups is illustrated in Figure 2.

The empirical overview demonstrates that influence potential is not a constant and varies across IOs. At the upper end of the graph are those organizations where member states have the largest potential for influence. For instance, both in the United Nations Population Fund (UNFPA) and United Nations Development Programme (UNDP), member states control all of the four political resources. In contrast, IPAs have the largest influence potential in IOs at the bottom end of the graph. Organizations like International Maritime Organization (IMO), United Nations Programme on HIV/AIDS (UNAIDS), United Nations High Commissioner for Refugees (UNHCR), and United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA) hence constitute extreme cases of IPA influence potential.[10] In the remaining cases in the middle of the graph, none of the stakeholder groups clearly dominates in terms of political resources. Yet, IPAs still tend to slightly dominate in five of the middle IOs, while member states have more political resources in United Nations Educational, Scientific and Cultural Organization (UNESCO), World Health Organization (WHO), United Nations Industrial Development Organization (UNIDO), and Food and Agriculture Organization (FAO). Influence potential in United Nations Office of Internal Oversight Services (UN OIOS) and International Atomic Energy Agency (IAEA) is equally shared by both member states and IPAs.[11] Interestingly, our findings suggest that in the majority of IOs (14 of 24), political resources for evaluation influence are dominated by IPAs. It also shows that IPAs tend to dominate informal

resources in more than half of the IOs, agenda-setting powers in 15 IOs, and budgetary political resources in even 20 organizations.

Based on the revealed variation, the 24 IOs can be classified into three groups. As mentioned above, the upper and lower ends constitute extreme cases of either member state or IPAs' influence potential. They can thus be treated as most likely cases for empirically studying the actual stakeholder impact. Conceptually, the identified political resources may be considered as first interaction points to analyze the politics of evaluation. UNDP, for instance, regularly organizes stakeholder meetings 'towards the end of the evaluation process for all IEO [evaluation unit's] evaluations' (UNEG, 2017b). Observing such events may help to analyze whether and how member states use these meetings as 'window of opportunity' for articulating their own evaluation-related preferences. Furthermore, the IOs in the middle of the graph constitute most likely cases to observe contestation or 'wars' in stakeholder influence. Since none of the stakeholder groups dominate the influence potential, diverging evaluation-related interests should result into contestation in the selection and conduct of evaluation. Hence, studying these IOs might help to compare the four categories of political resources and assess their *actual* use in the exercise of influence. For example, if one could observe that member states successfully affect evaluation results in UN OIOS (although IPAs dominate the other two political resources), one could assume that formal bases of influence are more useful than informal ones to translate own preferences.

Our findings also have implications for research on evaluation use (Patton, 1997; Picciotto, 2016; Weiss, 1993). Although we demonstrate variation in influence *potential*, we expect stakeholders to indeed use political resources at their disposal. Evaluation results should therefore reflect the preferences of more influential stakeholders, which, in turn, should also affect evaluation use (Weiss, 1993: 103). Thus, different levels and purposes of evaluation use can be expected in the three groups of IOs. For instance, if member states have accountability-related evaluation preferences in IOs of extreme member states potential (Figure 2), we should observe evaluations primarily used for accountability, even if IPAs prefer learning or legitimacy. Regarding contested middle cases, one could expect a reduced evaluation use, since none of the stakeholders dominate the process. Similarly, extreme cases of IPA influence potential should reflect the use of evaluation for IPAs' interests.

Our findings imply that we should not simply assume that member states—as political decision makers—always have the largest influence over IO policymaking processes. To the extent that IPA-controlled evaluation enters the process, either ex ante to confirm regulatory need or ex post to allow adjustment or termination, policies could also be influenced along the lines of IPA interests. Such a finding would be in line with IO studies depicting IPAs as autonomous policymaking entrepreneurs (Bauer et al., 2017; Eckhard and Ege, 2016; Ege, 2017; Knill et al., 2016, 2018) as well as studies highlighting in domestic settings the influence of bureaucrats over evaluators (Bjornholt and Larsen, 2014; Leeuw and Furubo, 2008; Weiss, 1993).

Concluding remarks

Building on the notion that interest contestation makes evaluation political, this article set out to develop a conceptual tool for the comparative analysis of stakeholder influence potential across 24 UN organizations. Following a resource-based approach, we distinguished between four categories of political resources that allow stakeholders to potentially affect evaluation results. These resources refer to formal or informal social interactions and can be used either

through direct or indirect participation during the different stages of the evaluation process. The theorized concept speaks to the broader debate on the politics of evaluation (Bjornholt and Larsen, 2014; Chelimsky, 1987; Weiss, 1993) and is intended to serve as a practical tool to study stakeholders' influence. This article specifically focused on IOs that provided us with a reasonably similar group of entities to treat them as belonging to the same concept (cf. Beach and Pedersen, 2018). This is novel to the study of evaluation, which has traditionally been hampered by idiosyncrasies of national political systems or individual prominent organizations such as the European Union (Hojlund, 2014; Pleger et al., 2016). Nevertheless, the concept bears applicability beyond IOs alone since key stakeholder groups and their political resources exist both at the national and international levels.

Empirically, our findings demonstrate that stakeholder potential to influence evaluation systems varies between IOs. As formulated in the previous chapter, the revealed patterns of three IO groups with specific influence manifestations suggest expectations on evaluation use in concrete organizations. The three clusters also serve as plausible starting points for future studies to specify the concept and gain new knowledge on the politics of evaluation in IOs in particular.

Theoretically, this study shows that IPAs—and not member states—possess most of the political resources in the majority of the IOs, which supports the literature on the bureaucratic influence, in general, and on IPA influence, in particular (Bauer et al., 2017; Biermann and Siebenhüner, 2009; Eckhard and Ege, 2016; Goetz and Patz, 2017; Hooghe and Marks, 2015; Jörgens et al., 2016; Knill et al., 2016, 2018; Trondal et al., 2010). Recent studies on evaluation in IOs speculate that evaluation in a 'complex multilateral structure is not merely a tool to assess the merit and worth of projects or programs, but also as a way to institutionalize roles, relationships and mandates' (Raimondo, 2018: 34; Rutkowski and Sparks, 2014). While we cannot directly observe such behavioral patterns, our findings show that when it comes to the political resources for such influence, bureaucracies indeed dominate.

Our results may also be of practical use for those actually involved in the evaluation processes. Knowing who has the largest potential to influence evaluation results—and especially through what kind of political resources—may help other stakeholders to adjust evaluation design or to treat its results accordingly. Taking this into account may help evaluators themselves to be more aware of outside influence and maintain standards of impartiality and objectivity: 'only with sensitivity to politics of evaluation research can the evaluator be as creative and strategically useful as [s]he should be' (Weiss, 1993: 94).

While the introduced concept of stakeholder influence potential offers a range of analytical strengths, we recognize that the study is more suggestive than conclusive, since we refer to influence *potential* rather than influence as a causal act. We have highlighted methodological limitations both in light of the reduced complexity of IO evaluation processes and in terms of the data used. Nevertheless, we consider this article as a promising avenue for future studies and we provide specific expectations to be tested empirically. On this basis, further contributions can investigate the actual stakeholder influence in a more qualitative manner and improve the concept by specifying its assumptions and triangulating data sources.

Acknowledgements

The authors would like to thank all participants of the workshop on Policy Evaluation in the EU which was held on 25-26 January 2018 at the University of Maastricht Campus Brussels, in particular Paul Stephenson and Frans Leeuw, for comments on an earlier version of the manuscript. The United Nations Joint Inspection Unit (JIU) kindly provided additional background information on their methodology

and proceedings. We also thank Alexa Lenz for helpful research assistance and two anonymous journal reviewers for their comments which helped improving the manuscript.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors would like to gratefully acknowledge funding by the German Research Foundation (DFG, Grant Number: EC 506/1-1).

Notes

1. For important exceptions, see the works by Hojlund (2014), Raimondo (2018), Schoenefeld and Jordan (2017), and Weaver (2010).
2. For instance, although only 60 professionals were working in the United Nations (UN) evaluation system in the 1980s, almost a thousand expert consultants are now working together with about 160 full-time professional evaluators. In just a 2-years period (2012–2013), they have produced 600 evaluation reports which cost over US$100 million (JIU, 2014b: 8).
3. The term *evaluation use* 'refers to the way in which an evaluation and information from the evaluation impacts the program that is being evaluated' (Alkin and Taut, 2002: 1).
4. While the stages themselves include additional steps at the micro level, this reflects ideal-typical stages of evaluation processes as usually depicted in IOs evaluation guidelines (see, for instance, The Office of Evaluation [OED], 2015; UNEG, 2016).
5. These central evaluation units can be either stand-alone offices or co-located with management or audit functions. Four UN organizations do not have centralized evaluation offices and are thus excluded from the study. The distinction is based on the JIU (2014a) terminology.
6. These are (in the order shown in Figure 2) : UNFPA (United Nations Population Fund), UNDP (United Nations Development Programme), WFP (World Food Programme), ILO (International Labour Organization), UNESCO (United Nations Educational, Scientific and Cultural Organization), WHO (World Health Organization), WIPO (World Intellectual Property Organization), UNIDO (United Nations Industrial Development Organization), FAO (Food and Agriculture Organization of the United Nations), UN OIOS (United Nations Office of Internal Oversight Services), IAEA (International Atomic Energy Agency), WMO (World Meteorological Organization), UN Women (United Nations Entity for Gender Equality and the Empowerment of Women), UNODC (United Nations Office on Drugs and Crime), UNICEF (United Nations Children's Fund), UNCTAD (United Nations Conference on Trade and Development), UNEP (United Nations Environment Programme), UNHABITAT (United Nations Human Settlements Programme), ICAO (International Civil Aviation Organization), ITC (International Trade Center), IMO (International Maritime Organization), UNAIDS (United Nations Programme on HIV/AIDS), UNHCR (Office of the United Nations High Commissioner for Refugees), UNRWA (United Nations Relief and Works Agency for Palestine Refugees in the Near East).
7. Other actors, such as interest groups, non-governmental organizations (NGOs), and individuals (beneficiaries), may also seek to access evaluators from the outside; yet, not only are their influence capabilities lower, but so is the demand to affect evaluation results. For instance, individual citizens as beneficiaries of the World Bank cannot directly influence its policies. An immediate responsiveness to individuals' input demands is not an institutionalized obligatory part of the Bank's operational methods.
8. For instance, the evaluation unit in the UNDP operates as a completely separate office, while in IAEA it is integrated with oversight functions.
9. As the JIU report (2014a) indicated, the maturity of evaluation processes in different IOs varies as well. While the varying maturity level of evaluation functions may affect stakeholders' ability to actually exercise influence, it does not change the existence of the four political resources (i.e. influence potential).

10. For example, the UN Evaluation Group has emphasized that evaluation function in the IMO is indeed 'only focused on fulfilling the instruction of the Secretary-General,' the head of the IPA (UNEG, 2017a).
11. To explain the revealed variation, further studies should consider the different IO mandates and operating environments.

ORCID iD

Steffen Eckhard https://orcid.org/0000-0002-5320-0730

References

Alkin MC and Taut SM (2002) Unbundling evaluation use. *Studies in Educational Evaluation* 29(1): 1–12.

Azzam T (2010) Evaluator responsiveness to stakeholders. *American Journal of Evaluation* 31(1): 45–65.

Banfield EC (1961) *Political Influence: A New Theory of Urban Politics*. New York: The Free Press.

Banner DK (1974) The politics of evaluation research. *OMEGA* 2(6): 763–74.

Bauer MW and Ege J (2017) A matter of will and action: The bureaucratic autonomy of international public administrations. In:Bauer MW, Knill C and Eckhard S (eds) *International Bureaucracy: Challenges and Lessons for Public Administration Research*. London: Palgrave Macmillan, 13–42.

Bauer MW, Knill C and Eckhard S (2017) *International Bureaucracy: Challenges and Lessons for Public Administration Research*. London: Palgrave Macmillan.

Beach D and Pedersen RB (2018) Selecting appropriate cases when tracing causal mechanisms. *Sociological Methods & Research* 47: 837–71.

Biermann F and Siebenhüner B (2009) *Managers of Global Change: The Influence of International Environmental Bureaucracies*. Cambridge, MA: The MIT Press.

Bjornholt B and Larsen F (2014) The politics of performance measurement: 'Evaluation use as mediator for politics.' *Evaluation* 20(4): 400–11.

Bohne E (2010) *The World Trade Organization: Institutional Development and Reform*. Basingstoke; New York: Palgrave Macmillan.

Brandon PR and Fukunaga LL (2014) The state of the empirical research literature on stakeholder involvement in program evaluation. *American Journal of Evaluation* 35(1): 26–44.

Chelimsky E (1987) The politics of program evaluation. *Society* 25: 24–32.

Chelimsky E and Shadish WR (1997) *Evaluation for the 21st Century: A Handbook*. Thousand Oaks, CA: SAGE.

Clark TN (1968) *Community Structure and Decision-Making: Comparative Analysis*. San Francisco, CA: Chandler Publishing Company.

Cohen DK (1972) Politics and research: Evaluation of social action programs in education. In:Weiss CH (ed.) *Evaluating Action Programs: Readings in Social Action and Education*. Boston, MA: Allyn & Bacon, 137–65.

Cox RW and Jacobson HK (1973) *The Anatomy of Influence: Decision Making in International Organization*. New Haven, CT; London: Yale University Press.

Cronbach LJ, Ambron SR, Dornbusch SM, et al. (1981) Toward reform of program evaluation. *Educational Evaluation and Policy Analysis* 3(6): 85–7.

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.

Dahler-Larsen P (2012) *The Evaluation Society*. Stanford, CA: Stanford University Press.

Datta L-e (2011) Politics and evaluation: More than methodology. *American Journal of Evaluation* 32(2): 273–94.

Davidson J (2005) *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*. Thousand Oaks, CA: SAGE.

Eckhard S and Dijkstra H (2017) Contested Implementation: The unilateral influence of member states on peace building policy in Kosovo. *Global Policy* 8(1): 102–12.

Eckhard S and Ege J (2016) International bureaucracies and their influence on policy-making: A review of empirical evidence. *Journal of European Public Policy* 23(7): 960–78.

Ege J (2017) Comparing the autonomy of international public administrations: An ideal-type approach. *Public Administration* 95(3): 555–70.

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.

Goetz KH and Patz R (2017) Resourcing international organizations: Resource diversification, organizational differentiation, and administrative governance. *Global Policy* 8(2): 5–14.

Graham ER (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): 366–90.

Hardt H (2013) Informal norms: Shaping behavior in international organizations. In:Reinalda B (ed.) *Routledge Handbook of International Organization*. New York: Routledge, 337–49.

Hawkins DG, Lake DA, Nielson DL, et al. (2006) *Delegation and Agency in International Organizations*. Cambridge: Cambridge University Press.

Heinrich CJ (2002) Outcomes-based performance management in the public sector: Implications for government accountability and effectiveness. *Public Administration Review* 62(6): 712–25.

Hojlund S (2014) Evaluation use in evaluation systems—The case of the European Commission. *Evaluation* 20(4): 428–46.

Hooghe L and Marks G (2015) Delegation and pooling in international organizations. *The Review of International Organizations* 10(3): 305–28.

House ER (1974) The politics of evaluation in higher education. *Journal of Higher Education* 45: 618–27.

House ER (2008) Blowback: Consequences of evaluation for evaluation. *American Journal of Evaluation* 29(4): 416–26.

Howlett M, Ramesh M and Perl A (2009) *Studying Public Policy: Policy Cycles and Policy Subsystems*. Oxford: Oxford University Press.

Jacob S, Speer S and Furubo J-E (2015) The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation* 21(1): 6–31.

JIU (2014a) *Analysis of the evaluation function in the United Nations system: United Nations Joint Inspection Unit*. JIU/REP/2014/6. Geneva. Available at: https://www.unjiu.org/sites/www.unjiu.org/files/jiu_document_files/products/en/reports-notes/JIU%20Products/JIU_REP_2014_6_English.pdf

JIU (2014b) Complementary information. In:*Analysis of the Evaluation Function in the United Nations System: United Nations Joint Inspection Unit* (JIU/REP/2014/6). Geneva. Available at: http://www.fao.org/3/a-mr629e.pdf

Jörgens H, Kolleck N and Saerbeck B (2016) Exploring the hidden influence of international treaty secretariats: Using social network analysis to analyse the Twitter debate on the 'Lima Work Programme on Gender.' *Journal of European Public Policy* 23(7): 979–98.

Knill C, Bayerlein L, Enkler J, et al. (2018) Bureaucratic influence and administrative styles in international organizations. *The Review of International Organizations*. Epub ahead of print 1 March. DOI: 10.1007/s11558-018-9303-x

Knill C, Eckhard S and Grohs S (2016) Administrative styles in the European Commission and the OSCE Secretariat: Striking similarities despite different organizational settings. *Journal of European Public Policy* 23(7): 1057–76.

Lee B (2006) Theories of evaluation. In:Stockmann R (ed.) *Evaluationsforschung: Grundlagen Und Ausgewählte Forschungsfelder*. New York: Springer, 137–76.

Leeuw FL and Furubo J-E (2008) Evaluation systems: What are they and why study them? *Evaluation* 14(2): 157–69.

Liese A and Weinlich S (2006) Die Rolle von Verwaltungsstäben internationaler Organisationen: Lücken, Tücken und Konturen eines (neuen) Forschungsgebiets. In:Bogumil J, Jann W and Nullmeier F (eds) *Politik Und Verwaltung: Politische Vierteljahresschrift*. Wiesbaden: Springer VS Verlag, 491–524.

Lipson C (1991) Why are some international agreements informal? *International Organization* 45(4): 495–538.

Lowery D (2013) Lobbying influence: Meaning, measurement and missing. *Interest Groups & Advocacy* 2(1): 1–26.

Malen B, Murphy MJ and Geary S (1988) The role of evaluation information in legislative decision making: A case study of a loose cannon on deck. *Theory into Practice* 27(2): 111–25.

March JG and Olsen JP (1989) *Rediscovering Institutions: The Organizational Basis of Politics*. New York: Free Press.

March JG and Olsen JP (1998) The institutional dynamics of international political orders. *International Organization* 52(4): 943–69.

Mills WC (1956) *The Power Elite*. Oxford: Oxford University Press.

Morris M and Clark B (2013) You want me to do what? Evaluators and the pressure to misrepresent findings. *American Journal of Evaluation* 34(1): 57–70.

Morris M and Cohn R (1993) Program evaluators and ethical challenges. *Evaluation Review* 17(6): 621–42.

O'Brien T, Payne S, Nolan M, et al. (2010) Unpacking the politics of evaluation: A dramaturgical analysis. *Evaluation* 16(4): 431–44.

OED (2015) *OED Evaluation Manual: Office of Evaluation*. Rome: Food and Agriculture Organization of the United Nations.

Orr SK (2010) Exploring stakeholder values and interests in evaluation. *American Journal of Evaluation* 31(4): 557–69.

Patton MQ (1997) *Utilization-focused Evaluation: The New Twentieth Century Text*. Thousand Oaks, CA: SAGE.

Picciotto R (2016) Evaluation and bureaucracy: The tricky rectangle. *Evaluation* 22(4): 424–34.

Pleger L, Sager F, Morris M, et al. (2016) 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): 315–28.

Raimondo E (2018) The power and dysfunctions of evaluation systems in international organizations. *Evaluation* 24(1): 26–41.

Roker D (2005) Evaluation projects aimed at supporting the parents of young people: 'I didn't learn anything new, but …'. In:Taylor D and Balloch S (eds) *The Politics of Evaluation: Participation and Policy Implementation*. Bristol: Policy Press, 119–31.

Rossi PH, Lipsey MW and Freeman HE (2004) *Evaluation: A Systematic Approach*. Thousand Oaks, CA: SAGE.

Rutkowski D and Sparks J (2014) The new scalar politics of evaluation: An emerging governance role for evaluation. *Evaluation* 20(4): 492–508.

Schoenefeld J and Jordan A (2017) Governing policy evaluation? Towards a new typology. *Evaluation* 23(3): 274–93.

Scott WR (2007) *Institutions and Organizations: Ideas and Interests*. Thousand Oaks, CA: SAGE.

Scriven M (1991) *Evaluation Thesaurus*. Newbury Park, CA: SAGE.

Simon HA (1953) Notes on the observation and measurement of power. *Journal of Politics* 15(4): 500–16.

Stockmann R, Meyer W and Schenke H (2011) Independence of Evaluations. *Zeitschrift Für Evaluation* 10(1): 39–67.

Stone RW (2011) *Controlling Institutions*. Cambridge: Cambridge University Press.

Taylor D and Balloch S (2005) *The Politics of Evaluation: Participation and Policy Implementation*. Bristol: Policy Press.
The LSE GV314 Group (2014) Evaluation under contract: Government pressure and the production of policy research. *Public Administration* 92(1): 224–39.
Trondal J, Marcussen M, Larsson T, et al. (2010) *Unpacking International Organisations: The Dynamics of Compound Bureaucracies*. Manchester: Manchester University Press.
United Nations Evaluation Group (UNEG) (2016) *Norms and Standards for Evaluation*. New York: UNEG.
United Nations Evaluation Group (UNEG) (2017a) *Evaluation in IMO: Evaluation Function*. International Maritime Organization. Available at: http://www.unevaluation.org/about/memberagencies/detail/8 (accessed 17 December 2017).
United Nations Evaluation Group (UNEG) (2017b) *Evaluation in UNDP: United Nations Development Programme* (Use of Evaluation). Available at: http://www.uneval.org/about/memberagencies/detail/21 (accessed 17 December 2017).
Urpelainen J (2012) Unilateral influence on international bureaucrats. *Journal of Conflict Resolution* 56(4): 704–35.
Vaubel R, Dreher A and Soylu U (2007) Staff growth in international organizations: A principal-agent problem? An empirical analysis. *Public Choice* 133(3): 275–95.
Weaver C (2010) The politics of performance evaluation: Independent evaluation at the international monetary fund. *Review of International Organizations* 5(3): 365–85.
Weiss CH (1970) The politicization of evaluation research. *Journal of Social Issues* 25: 57–68.
Weiss CH (1972) *Evaluating Action Programs: Readings in Social Action and Education*. Boston, MA: Allyn & Bacon.
Weiss CH (1993) Where politics and evaluation research meet. *Evaluation Practice* 14(1): 93–106.
Weiss CH (1999) The interface between evaluation and public policy. *Evaluation* 5(4): 468–86.
Weiss TG and Jordan RS (1976) Bureaucratic politics and the world food conference: The international policy process. *World Politics* 28(03): 422–39.
Wergin JF (1976) The evaluation of organizational policy making: A political model. *Review of Educational Research* 46(1): 75–115.
Wildavsky A (1972) The self-evaluating organization. *Public Administration Review* 32(5): 509–20.

Steffen Eckhard is assistant professor for Public Administration and Organization Theory at the University of Konstanz. His research deals with the influence of the administration on policy-making and -implementation in national and international public organizations with a focus on crisis management and peace and security governance. His current projects address the politics of evaluation in international organizations as well as hybrid organizations in local crisis management.

Vytautas Jankauskas is a research fellow at the University of Konstanz and a PhD candidate at the Ludwig-Maximilians-University Munich. In his research, he explores the role of trust in international organizations, the impact of evidence-based policy making at the international level, and the dynamics of cybersecurity governance.