Evaluation Capture, Evaluator Resilience, and the Need for Competencies of Evaluators

Frans L. Leeuw[1]
Maastricht University

Lyn E. Pleger
Zurich University of Applied Sciences

Background: Not applicable.

Purpose: Not applicable.

Setting: Not applicable.

Intervention: Not applicable.

Keywords:

Research Design: Not applicable.

Data Collection and Analysis: Not applicable.

Findings: Not applicable.

---

[1] Important parts of this paper come from Raimondo and Leeuw (2020) and Pleger and Leeuw (2021).

Evaluation Capture: What Is It?

"Evaluation capture" doesn't sound pretty. And rightly so. Compared to its older family member, regulatory capture,[2] the new kid on the block is still less well-known and visible. Regulatory capture is the tendency of regulators to identify with the interests of market organizations and other societal organizations they are supposed to regulate. This occurs when a public authority charged with oversight and regulation comes to identify the public interest with the interests of these specific organizations, rather than the interests of the general public. One example is government ministries that want to develop and implement solar energy technology, and so hire consultants to help them on a large number of issues that influence the government to create standards that give preference to their particular technology. Another example is a government department that is overly concerned by small but unusually vocal groups on social media, as opposed to delivering on their mandate. For instance, a city department that ignores or delays implementing the instructions of the mayor's office in order to avoid criticism on social media (Naimonet, n.d.).[3]

However, the underlying mechanisms are not limited exclusively to regulators or policy makers, but can also be applied to other settings and sectors. As Dal Bó (2006, p. 204) puts it, regulatory capture "is, in fact, common to other areas of public policy and, like many of the theory's developments, it may be applied to the problem of political influence as it is broadly understood." It is therefore appropriate to apply the theory of regulatory capture to evaluations.

Evaluation capture is the situation in which evaluations and evaluators are surrounded by protocols, guidelines, standards, norms, criteria, templates, oversight, and reviews from national and supranational organizations, governments, NGOs, and agencies that ask for and finance monitoring and evaluation (M&E) activities; these protocols and guidelines are sometimes labeled or part of "evaluation policies." It should be noted that, up until now, this issue of capture has mainly been studied in larger organizations (national and international). Examples include texts that specify which type of evaluation should be done ("theory-driven evaluation approach is necessary"; "follow the counterfactual approach"); when and how results will be published; what legal-contractual aspects of evaluations are; on which criteria/standards the knowledge products will be judged; etc.

Take the example of evaluation capacity development (ECD) and M&E systems. It is often suggested that developing and installing ECD is crucial to help produce data, to train people to do evaluation work, and, in line with that, to contribute to the effectiveness of programs and policies. However, there are several reasons to be concerned with implementing more and more ECD and (related) M&E systems in more and more places. Horton et al., in an almost-20-year-old report, "Evaluating Capacity Development," found that ECD had to deal with

> [producing] many reports that arrived too late, after decisions had already been taken. Even when evaluation reports were understandable, arrived on time and addressed important issues, decision-makers often seemed to ignore them... The limited use of evaluation results has come to be viewed as the Achilles heel of evaluation. (2003, pp. 107–108)

In an evaluation of the World Bank's results-based evaluation system, Clements et al. (2008, p. 209) highlight the dysfunction of this approach: "The disadvantage of results-based M&E is that it does not establish the worth of program results. A program that reaches all its timid targets may be less cost-effective than one that fails to reach ambitious goals." Indeed, a results-based evaluation regime establishes incentives for program planners to select targets that are easier to reach. Moreover, as highlighted by Bamberger, Vaessen, and Raimondo (2015), objectives-based evaluation systems have also failed to address and assess complex bureaucratic interventions, and notably to scrutinize the many unintended consequences of bureaucratic actions, which was one of the reasons why they were created in the first place.

---

[2] Harris (2015) coined "developmental capture" in his study of Thailand's universal health care policy: "a network of state bureaucrats who occupied executive positions in a number of different departments and organizations within the state. From these positions, members of this network drew on a number of different strategic resources in the political, bureaucratic, civil, and international realms to advance an inclusive developmental agenda and managed to institutionalize universal health care" p. 165). "Audience capture" is, again, a somewhat related concept: a self-reinforcing feedback loop that involves telling one's audience what they want to hear and getting rewarded for it.

[3] Handbooks of political science and of public choice economics give numerous other examples.

Vallejo and Wehn (2016) analyzed evaluation capacity development and summarized the strengths and weaknesses of this type of activity in the context of the world of development. They concluded:

> Current results-based evaluation methods may provide a good measurement of CD [capacity development] outputs and outcomes that can be captured by indicators to report progress to donors. However, they are not able to measure impact, as CD projects, restricted by previously agreed budgets, resources, and time frames, are usually not designed to evaluate the sustainability of change and its impact over the medium or long term. (p. 10)

Finally, the "law of diminishing returns" applies here. The economic theory predicts that after an optimal level of capacity, additional activities (in this case, more capacity), will decrease the output (in this case, of relevant and applicable information) and can possibly also lead to the production of unintended, negative side effects. Examples of such side effects mentioned in the literature include "cooking the data," bureaucratization, red tape, dramaturgical compliance, and the performance paradox (Natsios, 2011; Pollitt, 2013; van Thiel & Leeuw, 2002).

The concept "evaluation capture" is related to:

- the institutionalization of evaluation;
- evaluation as standard operating process/procedure (SOP) (Dahler-Larsen, 2018). Often, evaluation policies, regulations, and protocols tend to develop mechanistic, rigid rules that mandate "everything should be evaluated every x year(s)," creating an "evaluation machine." This contributes to what has been called "evaluitis" which leads to forgetting about the value added (if there is any) of more evaluations.
- the panacea problem of evaluation (a situation where a systematic regime requires evaluations to be carried out as a general prescription, even where they are not needed, just because otherwise it would not be systematic evaluation) (Dahler-Larsen, 2018);
- evaluation systems (permanent, predictable, formal sets of activities) (Leeuw & Furubo, 2008).

These developments are, seen on their own, successes in establishing an evaluation structure, but, in their togetherness and interconnectedness, they make evaluations become more and more captured. Separately, having protocols to which evaluation research (and the terms of reference the study has to comply with during—for example—competitive tendering) must comply, or carrying out such research in a planned and periodic manner, or having regulations for its methodology, if dosed in size, is not similar to "being captured." But, when organizations have these and other provisions and deploy them simultaneously, then there is a significant risk of capture.

Possible Underlying Mechanisms

Raimondo and Leeuw (2020) have discussed several (possible) mechanisms for evaluation capture. One is that the institutionalization of evaluation systems within bureaucracies has created the *wrong incentives for evaluators, evaluands, policy makers, and politicians*. The incentive is not to develop and deliver the best policies, but rather to have evaluation systems run to reduce reputational risks for principals, managers, and sponsors. When evaluation resides within the bureaucracy, gaming evaluations' timing or influencing their contents (Pleger & Sager, 2018) is more doable than it is when evaluation happens in a more independent way and outside the bureaucracy.

Related to this is that evaluators can act as *budget-maximizing agents*. The budget is the number of evaluations that can be done, the money and social rewards that can be won. Assuming that most (professional) evaluators are of the opinion that the more evaluations they can do or participate in, the better it is (for them and for policy makers—and maybe also for beneficiaries and target groups), this opinion will drive their behavior. What Bamberger et al. (2004) called "shoe-string" evaluations, which can imply pinching evaluators, may stimulate this behavior. Some may decide to drop out, but others will be keen to get new contracts under better financial conditions.

A third mechanism deals with the *politics of evaluation systems in bureaucracy*. Several authors have questioned the assumption that an evaluation system is a politically neutral instrument initiated by principals to steer implementing agents, instead claiming that evaluation systems also steer principals and what is politically achievable (e.g., Bjørnholt & Larsen, 2014; Weiss, 1973). Performance measurement and evaluation are presented as instruments of governance. Weiss (1973) was among the first to explicitly present evaluation as an eminently political exercise. An

additional political dimension of evaluation relates to the role that key organizations, such as the Organisation for Economic Co-operation and Development and the World Bank, have played in promoting a global agenda for evaluation and the universalizing of evaluation standards and criteria. Evaluation is thus increasingly positioned within a global governance strategy that seeks greater influence for bureaucracies (Rutkowski & Sparks, 2014).

Raimondo (2016, 2018) identified a range of additional mechanisms that explain the bureaucratic capture of evaluation systems within international organizations. One is the way in which "loose coupling" is handled (by evaluators). Building on organizational sociology and on theories of international organization culture developed by Barnett and Finnemore (2004), Raimondo shows that evaluation systems in bureaucracies find their roots in the willingness of internal and external principals to remedy organizational loose coupling—the gaps between objectives and implementation, discourse and reality. Evaluation systems within bureaucracies can become powerful and legitimate by manifesting their supposed functional and structural independence, neutrality, and scientific and apolitical judgment on programs' worth. Actors operating in the name of a "results-based decision-making process" seek to deploy relevant knowledge to determine the worth of organizational projects and, indirectly, of the organization and its staff.

Ultimately, evaluation criteria may become the new organizational goals (Dahler-Larsen, 2012, p. 80), and new rules about how goals ought to be pursued are set.

What to Do to Remedy Capture? The Role of the Evaluator's Resilience

Based on the preceding remarks, it can be assumed that evaluators are not willing or able to close their eyes to capture as an unintended, adverse (if not perverse) (side-)effect of a 60-year-old "battle" to get evaluation on the agenda of public, semi-public, and private organizations and bureaucracies. Therefore, evaluators must deal with the capture issue in such a way that they do not throw the champagne (i.e., all the good and relevant characteristics of M&E) out with the cork. This implies that evaluators need to have certain competencies.

The first competence is to diagnose situations in which M&E have become, or are becoming, standard operating procedures (SOPs) or even rituals.[4] In evaluation training courses, attention should be paid to doing such a diagnosis, starting with being able to notice the very beginnings of SOP-ification. In Table 1, we give some examples of processes and situations that might act as alerts and indicators of capture; these examples could be part of such training.

---

[4] More than two decades ago, Mike Power published his book *The Audit Society: Rituals of Verification* (1997), which Leeuw (2010) has used in analyzing costs and benefits of evaluation.

Table 1. Processes and Situations Indicating Capture

Evidence-Based Decision Making
Asking for and doing implementation studies without taking into account the results of thousands of earlier implementation studies of archetypes of interventions, programs, and policies. Archetypical interventions and programs use classical mechanisms such as sticks, carrots, pillories, sermons, fear arousal, etc. Although the exact designs of the interventions may differ, what is under the hood is often the same as before (i.e., sticks, carrots, etc.). Statements about the chances of implementation failure (or success) can be made based on the results of previous implementation evaluations. Doing that is crucial before launching a new study. Commissioners who have evaluation money to spend are (sometimes) reluctant to follow this line of thought, as it may lead to critical questions higher up in the organization about why "the money was not spent on the commissioning of a new evaluation." Well, the answer is that when the intervention is "old wine in new bottles," it is largely symbolic to start with a new empirical study without having learned from existing work.[5] As, certainly in Western countries, an important industry has emerged (the behavioral program / intervention factories), the likelihood that so-called "new" policy interventions / programs in fact are "old wine in new bags" is greater than ever before.

Theory of Change
Asking, if not requiring, a program (in its terms of reference or in other ways) to present a theory of change, even when it makes no sense to produce it or it is simply not possible to do. If in such a situation a theory of change is produced, it is usually either a simple logic model referring to input, throughput, and output, or a diagram with colors, lines, arrows, circles and artistic images, without (much) informative content. Claiming to develop or (re)construct a theory of change and next claim to test it, while in reality the theory of change is already believed or "verified" beforehand.

SOP-ification
Asking for more SOP-ification (producing more standard operating procedures for more and more issues) and presenting (digital) tools to do so. Instead of having evaluators and commissioners reflect and think through what has to be done in certain situations, the easy way is to apply SOPs from cookbooks or websites. Moreover, these SOPs function as a protective belt for commissioners and evaluators, in the sense that paying attention to cookbooks helps to create "assurance" in the research practice.

Diagnosis is not enough—learning and knowing what to do to help combat "capture" situations, instead of becoming part of them, is crucial. Here, the evaluator's *resilience* marches in. In general, resilience can be defined as "the capacity of a system to absorb disturbance and still retain its basic function and structure" (Walker & Salt, 2012, p. xiii). Evaluator resilience can be understood as the sum of different individual and personal competencies and can be fostered or hindered by institutional arrangements. More precisely, as we have defined it elsewhere, evaluator resilience is "evaluators' willingness and capability to resist/withstand destructive external influences with no closure to constructive influences by assertively following scientific requirements in order to produce … credible, valid and useful outcomes" (Pleger & Leeuw, 2021, p. 150). We have also distinguished four individual dimensions and four institutional factors capturing evaluator resilience;[6] these are shown in Figure 1 (Pleger and Leeuw, 2021, p. 151153). The institutional factors comprise the following: code of ethics, training/professionalism, work environment and structural independence, and evaluation management. The individual factors can be summarized as follows:

The first individual dimension comprises the character traits of the evaluator. Examples of

[5] For a more in-depth discussion of the issue of dynamics between commissioners and evaluators, see Pleger and Hadorn (2018).
[6] See Eckhard and Jankauskas (2019) for a more detailed discussion on the topic. They argue that evaluator resilience is not only a result of individual ability, but also (potentially) is influenced to some degree impacted by the institutionalization process—such as stakeholder influence potential.

personal traits that contribute to resilience are: good communication skills, emotional intelligence and being ethical. Secondly, a resilient evaluator requires methodological skills. That means they must have knowledge of a diversity of methods and be confident in the methodological approaches being employed. Beside these hard skills, evaluation resilience includes soft skills as the third dimension. Examples of these soft skills are relationship building, stress management, conflict resolution, communication skills, flexibility, adaptability as well as pragmatism. Finally, a resilient evaluator should have personal and professional integrity as a fourth individual dimension. Integrity refers to an "ethical backbone" and can be further described as being confident, not being compromised by clients due to livelihood considerations, independence of thought or professional self-confidence (Pleger & Leeuw 2021, p. 151–153; see Figure 1).[7]

Figure 1. Model of Evaluator Resilience
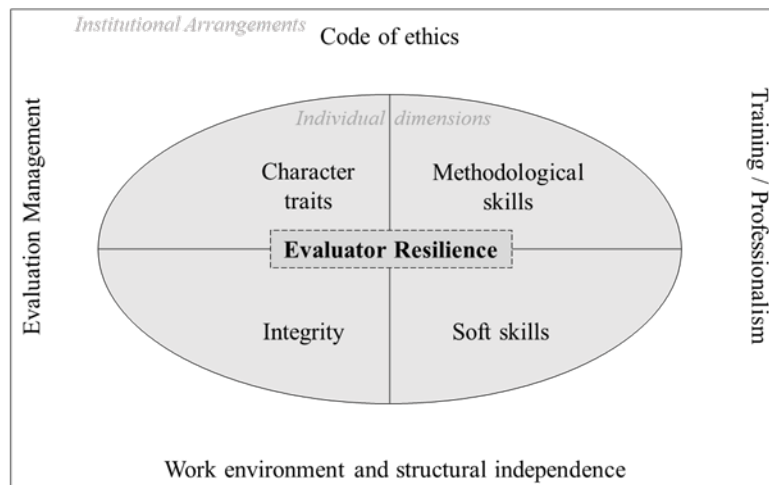
Note. From "Resilient Evaluators: Characteristics, Conditions and Prospects," by L. E. Pleger and F. L. Leeuw, 2021. In *Ethics for Evaluation: Beyond "Doing No Harm" to "Tackling Bad" and "Doing Good"* (p. 151), by R. D. van den Berg, P. Hawkins, and N. Stame (Eds.), Routledge.

Evaluator resilience encompasses not only a cognitive-intellectual dimension, but also an in-depth ethical-behavioral approach/style of evaluators, which may help to realize goals such as:

- collaborating with stakeholders/ commissioners while also fundamentally challenging them;
- being able to navigate between the demands of evaluation clients and the needs of a valid and credible scientific perspective in the evaluation;
- knowing how to deal with fake news, cancel culture, and "fake handbags" (e.g., selling logframes / logic models for intervention theories or theories of change).

In this world, evaluators must survive, produce relevant results, and avoid being sucked into the system of ever more and similar monitoring and evaluation activities. This issue is different from what evaluators faced 60 years ago, when it was a battle merely to get evaluations going; today's context requires resilient evaluators who are capable of the following:

- saying no to invitations to do evaluations if they are largely and only done for bureaucratic or symbolic reasons, as Vähämäki and Verger (2019) show;

---

[7] For a more in-depth discussion of specific evaluator competencies that these dimensions cover, see Pleger and Leeuw (2021).

- reformulating/rearranging evaluation questions and methods in such a way that substantive questions are addressed—questions that "speak truth to power"—while refraining from speech acts that constitute "evaluation machines";

- resisting being sucked into what Admiral Rickover once described as "say-do policies" (see Figure 2 below).

Figure 2. Say-Do

"In January 1982, Admiral Hyman G. Rickover, approaching retirement as director of the U.S. Navy's nuclear reactors and as irascible as ever, appeared before Congress for the last time. His testimony was full of blunt observations, none as cutting as his take on certain public servants. "I have an expression—the 'Say-Do,'" the 82-year-old Rickover declared. "People say something and the newspapers laud them, before they've done a damn thing and then they never do it and then they go on to some other Say-Do thing, and they get more credit." Detested by some, idolized by others, Rickover was, without a doubt, a doer. And he had little time for those who weren't" (Cole, 2015, para. 1).

Note. From U.S. Naval Historical ed States Naval History and Heritage Command) via Wikimedia Commons. In the public domain.

Final Thoughts

The overarching question this paper addresses is the nature of evaluation capture and how evaluators can remedy it. The answer lies in the presence of evaluator resilience. However, evaluator resilience is a complex concept. It requires manifold individual competencies of the evaluator, as well as certain institutional arrangements that enable the unfolding of these individual characteristics. Examples of such institutional arrangements include formal evaluation protocols and codes of how to act in stressful situations or moral education/training, but also independence, organizationally and personally as well as from upper management (Pleger & Leeuw, 2021). These are not easy conditions to create, and they ask quite something from evaluators. Evaluators need willpower to criticize the behavior just described, and to do so in socially skillful ways, understanding causes and suggesting alternative ways to move forward. This implies that a resilient evaluator not only knows his or her own emotions and uncertainties, including physical reactions, but also those of the persons he or she has to work with.

As stress is or will be an important ingredient of being resilient, neurofeedback on one's own stress levels may help. Insights from a relatively new transdiscipline (social neurosciences) can offer some guidance as evidence on emotion regulation and other related processes is produced (Pleger & Leeuw, 2021). Emotion regulation is a complex process involving cognitions and the functioning of parts of the brain (such as the amygdala; Ling et al., 2019). The progress in social neurosciences research includes experiments with hormones, use of face recognition software, studies with virtual reality and augmented reality to simulate stressful events and reactions of other persons, and

development and testing of new neurobiological interventions (focusing on decision-making under stress).[8] It is our hope (and belief) that using and developing these types of approaches and competencies when conducting evaluations can contribute to preventing or limiting capture as much as possible. Hopes and beliefs, however, need to be tested. So, time will tell.

References

Bamberger, M., Rugh, J., Church, M., & Fort, L. (2004). Shoestring evaluation: Designing impact evaluations under budget, time and data constraints. *American Journal of Evaluation*, *25*(1), 5–37. https://doi.org/10.1177/10982140040250010 2

Bamberger, M., Vaessen, J., & Raimondo, E. (Eds.). (2015). *Dealing with complexity in development evaluation*: *A practical approach*. Sage Publications. https://doi.org/10.4135/9781483399935

Barnett, M. N., & Finnemore M. (2004). *Rules for the world: International organizations in world politics*. Cornell University Press.

Bjørnholt, 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

Clements, P., Chianca, T., & Sasaki, R. (2008). Reducing world poverty by improving evaluation of development aid. *American Journal of Evaluation*, *29*(2), 195–214. https://doi.org/10.1177/1098214008318657

Cole, R. L. (2015, January 2). The admiral of the atom. *City Journal*. https://www.city-journal.org/html/admiral-atom-9847.html

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

Dahler-Larsen, P. (2018). The skeptical turn in evaluation. In J. E. Furubo & N. Stame (Eds.), *The evaluation enterprise* (pp. 58–80). Routledge. https://doi.org/10.4324/9780429470202-3

Dal Bó, E. (2006). Regulatory capture: A review. *Oxford Review of Economic Policy*, *22*(2), 203–225. https://doi.org/10.1093/oxrep/grj013

Eckhard, S., & Jankauskas, V. (2019). The politics of evaluation in international organizations: A comparative study of stakeholder influence potential. *Evaluation*, *25*(1), 62–79. https://doi.org/10.1177/1356389018803967

Harris, J. (2015). "Developmental capture" of the state: explaining Thailand's universal coverage policy. *Journal of Health Politics, Policy and Law*, *40(1)*, 165–193. https://doi.org/10.1215/03616878-2854689

Horton, D., Alexaki, A., Bennett-Lartey, S., Brice, K. N., Campilan, D., Carden, F., de Souza Silva, J., Duong, L. T., Khadar, I., Maestrey Boza, A., Kayes Muniruzzaman, I., Perez, J., Somarriba Chang, M., Vernooy, R., & Watts, J. (2003). *Evaluating capacity development: Experiences from research and development organizations around the world*. International Service for National Agricultural Research (ISNAR), the Netherlands; International Development Research Centre (IDRC), Canada; & ACP-EU Technical Centre for Agricultural and Rural Cooperation (CTA), the Netherlands. https://usingevidence.com/wp-content/uploads/2018/05/IDL-33404.pdf

Leeuw, F. L. (2010). Benefits and costs of evaluation—an essay. *Zeitschrift fur Evaluation*, *9*(2), 211–229.

Leeuw, F., & Furubo, J. E. (2008). Evaluation systems: What are they and why study them? *Evaluation*, *14* (2), 157–169.

Ling, S., Umbach, R., & Raine, A. (2019). Biological explanations of criminal behavior. *Psychology, Crime & Law*, *25*(6), 626–640. https://doi.org/10.1080/1068316x.2019.15727 53

Naimonet. (n.d.). *4 examples of regulatory capture*. https://naimonet.com/en/regulatory-capture

Natsios, A. (2011). The clash of the counter-bureaucracy and development. *Center for Global Development*. https://www.cgdev.org/publication/clash-counter-bureaucracy-and-development

Naval History and Heritage Command (United States). (2023, January 26). *Hyman George Rickover, 27 January 1900 – 8 July 1986*. https://www.history.navy.mil/content/history

---

[8] The former head of evaluation of the UNDP (Indran A. Naidoo) had a U.S. neuroscientist work with his evaluation team in New York while using insights that Dr. S. Pillay shared with them; in his new position as head of evaluation in IFAD (International Fund for Agricultural Development), he continues to do so. Leeuw has been codirector of the National Program on Brain and Cognition Studies of the Netherlands National Science Foundation, with the goal to make insights from cognitive/social neurosciences relevant to the policy practice and to evaluation.

/nhhc/research/library/research-guides/modern-biographical-files-ndl/modern-bios-r/rickover-hyman-g.html

Pleger, L. E., & Hadorn, S. (2018). The big bad wolf's view: The evaluation clients' perspectives on independence of evaluations. *Evaluation*, *24*(4), 456–474. https://doi.org/10.1177/1356389018796004

Pleger, L. E., & Leeuw, F. L. (2021). Resilient evaluators: Characteristics, conditions and prospects. In R. D. van den Berg, P. Hawkins, & N. Stame (Eds.), *Ethics for evaluation: Beyond "doing no harm" to "tackling bad" and "doing good"* (pp. 141–162). Routledge. https://doi.org/10.4324/9781003247234-9

Pleger, L. E., & Sager, F. (2018). Betterment, undermining, support and distortion: A heuristic model for the analysis of pressure on evaluators. *Evaluation and Program Planning*, *69*, 166–172. https://doi.org/10.1016/j.evalprogplan.2016.09.002

Pollitt, C. (2013). The logics of performance management. *Evaluation*, *19(4)*, 346–363. https://doi.org/10.1177/1356389013505040

Power, M. (1997). *The audit society: Rituals of verification*. Oxford University Press.

Raimondo, E. (2016). *The institutionalization of monitoring and evaluation systems within international organizations: A mixed-methods study* [Doctoral dissertation, George Washington University]. ProQuest Dissertation Publishing.

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

Raimondo, E., & Leeuw, F. L. (2020). Evaluation systems and bureaucratic capture: Locked. In B. Perin & T. Tyrell (Eds.), *Changing bureaucracies: Adapting to uncertainty, and how evaluation can help* (pp. 143–160). Routledge.

Rutkowski, D., & Sparks, J. (2014). The new scalar politics of evaluation: An emerging governance role for evaluation. *Evaluation*, *20*(4), 492–508. https://doi.org/10.1177/1356389014550561

Vähämäki, J., & Verger, C. (2019). *Learning from results-based management evaluations and reviews* (OECD Development Co-operation Working Papers, 53). Organisation for Economic Co-operation and Development.

Vallejo, B., & Wehn, U. (2016). Capacity development evaluation: The challenge of the results agenda and measuring return on investment in the global South. *World Development*, *79*, 1–13. https://doi.org/10.1016/j.worlddev.2015.10.044

van Thiel, S., & Leeuw, F. L. (2002). The performance paradox in the public sector. *Management Review*, 25(3), 267–281. https://doi.org/10.2307/3381236

Walker, B., & Salt, D. (2012). *Resilience thinking: Sustaining ecosystems and people in a changing world*. Island Press.

Weiss, C. H. (1973). Where politics and evaluation research meet. *Evaluation*, *1*(3), 37–45. https://doi.org/10.1016/0886-1633(93)90046-R