Department for Evaluation

EVAL Reflection Note series – no. 2

Independence: When, why and how it matters for evaluation credibility

October 2024

This note has been written by Ida Lindkvist PhD , senior advisor at the Department for Evaluation at Norad and Megan Kennedy-Chouane, head of Evaluation Unit and lead of the DAC Network on Development Evaluation at the OECD.

Suggested citation: Lindkvist, I., & Kennedy-Chouane, M. (2024). *Independence: When, why and how it matters for evaluation credibility* (No. 2; EVAL Reflection Note series). Evaluation Department, Norad.

ISBN: 978-82-8369-211-2
Cover photo: Nadia Frantsen

1. Introduction

High-quality evaluations provide credible and useful information for decision-making. This reflection note discusses the importance of independence for both credibility and usefulness, and how independence can be ensured, under different institutional settings. By independence we refer to an organisational structure for evaluation where the evaluation function is distinct from those that have a stake in evaluation results, and where the function has full authority for what to evaluate and how. Once the evaluation is commissioned, the structure provides evaluators, which could be internal or external to the evaluation function, with full authority to carry out the evaluation within the terms of reference for that evaluation.

The purpose of the note is to clarify when independence is required to ensure evaluation credibility. The key message is that independence is needed to ensure credibility if those that commission or conduct the evaluation are perceived to have a stake in specific evaluation findings or recommendations, and if the evaluation is to be used by a different stakeholder group for accountability purposes.

If independence is required, we argue that both commissioners and evaluators need to be independent. This is because perceived bias may arise both during the decision phase of what and when to evaluate, and when the evaluation is carried out.

The reflection note will focus on the role of credibility and independence, and will not, or only to a limited extent, address other key issues that are important for evaluation quality, such as ethics, usefulness, the evaluation process and so forth.

2. Evaluation purpose

It is generally acknowledged that the purpose of evaluation is to provide evidence that can be used for accountability and learning, and that evaluations need to be of high quality – i.e., both credible and useful to serve this purpose.

The OECD's Development Assistance Committee's evaluation principles (OECD/DAC, 1991, p. 5) set out the purpose of evaluation as follows: "§6 (...): - to improve future aid policy, programmes and projects through feedback of lessons learned; and - to provide a basis for accountability, including the provision of information to the public."

Accountability in this context means that information about results should be provided to the public so that it can be used in both donor and partner countries[1]. The principles specifically state that evaluation accountability in this context is distinctly different to legal or audit accountability.

Although most agree that the purpose of evaluation is accountability and learning[2], there is less agreement as to what credibility and usefulness mean and how this should be achieved. We argue that independence should be considered on a spectrum, and as a factor of both use and credibility, rather than a simplistic dichotomy of independent = good and not independent = bad.

3. Credibility and use – Key determinants of evaluation quality

High-quality evaluations are both credible and useful. Before delving into the concept of credibility, let us briefly consider evaluation use.

According to Kupiec et al. (2023), we can differentiate between three different types of use: a) instrumental and conceptual, b) symbolic and ritualistic use, and c) regulatory use. Instrumental use refers to the use of evaluative evidence for decision-making, while conceptual use refers to evidence that is produced with the purpose of "influencing the minds of decision makers without any immediate impact on their decisions" (Kupiec et al., 2023, p. 340). Symbolic and ritualistic use, on the other hand, refers to evaluations that are undertaken for symbolic or ritualistic reasons, not because they are needed for any specific purpose apart from serving as a symbol or being part of a ritualistic process. Examples of regulatory uses could be the decision to initiate an evaluation for contractual reasons – for example, an agreement requiring a mid-term or end review. Other types of use also

---

[1] "§8 The accountability notion of evaluation referred to here relates to the development results and impact of development assistance. It is distinct from accountability for the use of public funds in an accounting and legal sense, responsibility for the latter usually assigned to an audit institution. Information about the results of development assistance should be provided to the public and their leaders in both donor and recipient countries" (OECD/DAC, 1991, p. 5).

[2] Today some argue that evaluations going too far in the direction of accountability (and audit) will prevent learning from taking place as the evaluation will be more focused on defending against critique rather than critically adjusting its course of action. Others argue that it is not useful to use evaluations for learning unless they also provide a truthful and credible account of results. See Reinertsen et al. (2022) for a discussion.

exist, and different types of use can be combined. We will not get into these here – however, we refer interested readers to Vedung (2012).

What is important to keep in mind is that evaluations may have different user-groups and may therefore also have to satisfy different types of uses at the same time. For example, an independent evaluation function may carry out an evaluation that serves both instrumental and regulatory use, in the sense that the independent function is mandated to provide information to the public while, at the same time, the evaluation may, for example, provide a programming unit with information that it can directly use to improve its work.

Credible evaluations can be trusted and are trusted by key stakeholders. By stakeholder we mean anyone who can be affected by changes in the policies, projects or programmes evaluated, and/or anyone who takes an interest in the outcome of the policy, project or programme. In development assistance, typical stakeholder groups could be a bilateral or a multilateral funder, a programming unit in a partner agency (i.e., those that design and/or implement programmes), beneficiaries in partner countries, taxpayers in the country from which the funding originates, or other civil society organisations based anywhere.

That they 'can be trusted' means that evaluations are unbiased, to the extent that is possible, and hold high methodological quality. By unbiased we mean that the choice of questions and approach to the evaluation is not (intentionally) biased to favour one stakeholder group – for example, deciding not to ask questions that will portray a stakeholder in a bad light. Methodological quality ensures that the responses to evaluation questions are valid and reliable, whilst unbiased here refers to the evaluation not being altered due to undue pressure from stakeholders[3].

We follow Kupiec et al. (2023) and make a distinction between influence and pressure. Pressure refers to certain stakeholder groups using their position to bias the evaluation, whilst influence is positive and refers to stakeholders engaging in activities that positively affect the evaluation's ability to fulfil its purpose. A stakeholder that is facing an incentive structure that would mean that their interests would be better served if specific questions were not asked, or if they were asked would be better off with a specific evaluation finding, could apply pressure either when the evaluation is developed or when it is carried out. Influence, on the other hand, would mean that they use their resources, for example, to aid in the access to data or similar that could improve the quality of the evaluation.

With respect to methodological quality, relevant data must be available, and the methodology used for data collection and analysis needs to be appropriate for the purpose and well applied[4].

That evaluations 'are trusted' means that key stakeholder groups *perceive* the evaluation to be unbiased.

---

[3] Bias can, of course, also endanger methodological quality; in this discussion paper, we are mainly concerned with bias arising due to stakeholder pressure.

[4] Although methodological quality is certainly key for credibility and remains a key challenge for many donor agencies, this is not a focus here. See, for example, OECD (2023), and for a full overview of quality elements, see *Quality Standards for Development Evaluation* (OECD, 2010).

Note that low credibility due to perceived bias can occur without any threat to actual methodological quality. Analysis of data may provide methodologically reliable and valid conclusions but may not be perceived to be credible because specific stakeholders do not trust an evaluation process where the evaluation is commissioned or carried out by a stakeholder whose incentive structures mean that they would be better off with a specific result.

The main point is that it is not sufficient that an evaluation upholds high methodological quality and is free from undue pressure from stakeholders if it is not perceived to be credible. If key stakeholders do not trust the findings, they are unlikely to use them.

4. Incentive structures in development assistance

Evaluations are conducted for a specific use (often related to the general purposes of learning and accountability) and can have direct funding and policy implications. This means that evaluations may have real and sometimes immediate consequences for stakeholders, and that those who are affected can have an incentive to exert pressure on the evaluation in a specific manner to safeguard their interests[5].

What makes evaluations of development co-operation different from evaluations conducted solely within a national context is that funders are, by definition, based outside the country in which an intervention takes place. Furthermore, the partners carrying out interventions may or may not be locally based, and may or may not be part of a national public system. This means that the taxpayers who are paying for development assistance are not directly affected by what they fund, and that those who are directly affected may have limited influence over those implementing the interventions, or over funders.

The latter is sometimes called a broken feedback loop because of the missing link between funders and beneficiaries (Lindkvist & Dixon, 2014). Unless taxpayers[6] are provided with information about results, funders cannot be held accountable for ineffective or inefficient spending. The inherent power dynamics and information imbalances in international co-operation can also make it easier for some stakeholder groups to exert pressure in order to bias the evaluation.

The development of an "aid industry" in itself has also raised concerns, as partners receiving funding may present "glossy brochures" with only positive results, and donors and implementers are believed to care more about the size of budgets than results (see, for example, *The Cartel of Good Intentions* by Easterly (2002)).

---

[5] That said, the line between commissioned research and evaluation may sometimes be unclear.

[6] Information about results can also be used by beneficiaries of aid to hold service providers accountable, as has been described by Svensson & Bjorkman (2007).

Suppression of negative results can occur even when partners or donor agencies genuinely care about results – for example, if they care more about the objectives of specific development assistance programmes than taxpayers in donor countries. Taxpayers may be less willing to accept mistakes, compared with well-meaning development workers who may wish to avoid negative results as these could stifle funding (Pritchett, 2002). That said, the idea that bureaucratic agencies maximise budgets is not new and does not only apply solely to the field of development assistance (see, for example, Niskanen (1968)).

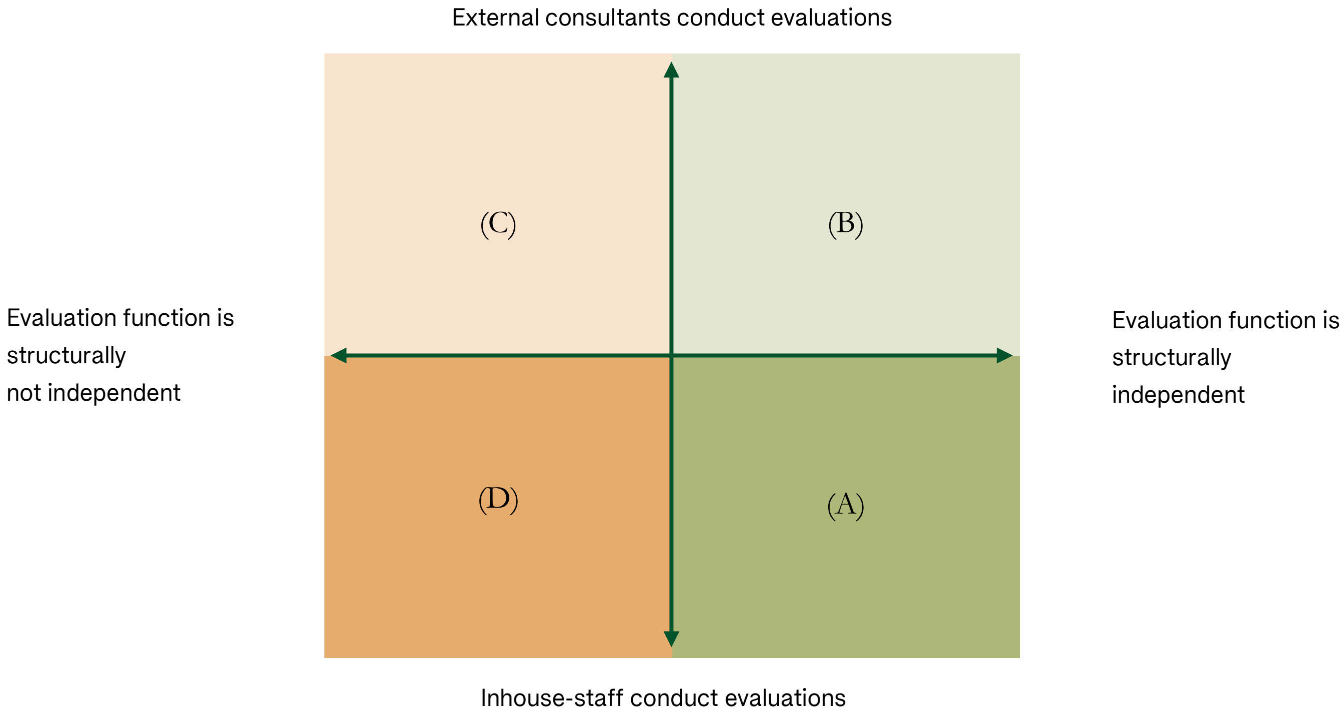
5. Independence and different institutional arrangements

The concern about stakeholder pressure in international development assistance could possibly explain why the OECD's Development Assistance Committee has introduced independence as a key principle for development assistance (OECD/DAC, 1991, p. 5): "The evaluation process should be impartial and independent from the process concerned with policymaking, and the delivery and management of development assistance." The principles specify that this means a separation between those that develop and carry out the evaluation and those that are evaluated[7], and that the entire evaluation process, from deciding what to evaluate and how, should be independent. Again, in this context, independence is a means of avoiding a specific type of bias – namely, untoward pressure from self-interested stakeholders in the intervention – and is not an end goal in itself. However, other biases and blind spots, including any (implicit) biases of external evaluators, may persist.

[7] "§16. Impartiality and independence will best be achieved by separating the evaluation function from the line management responsible for planning and managing development assistance. This could be accomplished by having a central unit responsible for evaluation reporting directly to the minister or the agency head responsible for development assistance, or to a board of directors or governors of the institution" (OECD/DAC, 1991, p. 6).

Although many different constellations exist for bilateral development actors, the evaluation function is normally either based outside or inside the development agency. Common constellations are depicted in Figure 1 below.

Figure 1 Different ways to organise the evaluation function



Various institutional arrangements can be used to ensure independence, including budget and reporting lines, autonomy in deciding what and when to evaluate, or the establishment of procedures to appoint the evaluation director. The degree of institutional or structural independence of the evaluation function is depicted on the x-axis in Figure 1.

Another difference is the extent to which the evaluation function's staff carry out evaluations, or whether evaluations are conducted by external consultants. This is depicted on the y-axis in Figure 1. Some evaluation functions conduct evaluations themselves, whilst others do not, or do both – i.e., fully outsource some to others and conduct some themselves. The latter has been the case for the independent Department for Evaluation at Norad. In Figure 1, A and B refer to independent evaluation functions that either commission independent evaluations, i.e. appoint external consultants to conduct the evaluation (B) or carry out independent evaluations relying on own staff (A).

Bilateral donors with an independent evaluation function normally also see the need to carry out their own evaluations, or to fund or commission evaluations. These evaluations are sometimes called decentralised evaluations, self-evaluations or similar. Hereafter, we will refer to these as self-evaluations. If they are carried out by external consultants, they fall under category C, and if they are carried out in-house by their own staff, they fall under category D.

In Figure 1, C – and especially D – would not be regarded as independent according to the OECD/DAC principles; however, these evaluation types could still provide credible evaluation findings for specific types of use. We will return to this in the discussion below.

6. Discussion of different models

The strictest form of institutional independence is to place the evaluation function outside the organisation whose work is to be evaluated, thereby cutting off the evaluation function from internal documents and day-to-day meetings with internal staff. This can potentially be, or be perceived to be, more independent, although there is a risk that the greater distance between the evaluation function and the organisation's work could influence the relevance of evaluations, and thereby their use. Organisational cultures and procedures also create a variety of incentives for staff and external actors, including implementing partners and consultants. For instance, many development agencies have found that spending pressure (hitting disbursement or financial targets) is the overriding incentive, which can undermine staff's efforts to reflect and learn through evaluation and may also create incentives to downplay bad news (see, for example, the evaluation *Can we Demonstrate the Difference that Norwegian Aid Makes?*, which argued that the organisation was geared towards spending rather than results (Department for Evaluation, 2014)).

Given the importance of both "actual" and perceived credibility, and the role that independence plays in each, institutional independence is likely to be required when one stakeholder group is to be provided with information from evaluations in order to hold another accountable. If we were to choose model D, the stakeholder would both commission and carry out the evaluation, which a second stakeholder would then use to hold the first stakeholder accountable. Model D would be unlikely to result in credible evaluations in this instance. However, model C might also be problematic when evaluations are to be used for accountability purposes. According to Michaelowa & Borrmann (2006), the actual impendence of independent consultants would depend on the market for evaluations.

Michaelowa and Borrmann's reasoning is that the market for evaluations will determine the real independence of consultants. If consultants rely on the same commissioner to get work, they may have to adopt the preferences of the commissioner to secure future work. For example, if an agency has a stake in a specific finding, consultants are likely to be aware of this. If they also know that their portrayal of those findings will influence their opportunity to get work from the same funder, they then have an incentive to look for information that supports that finding[8]. In addition, even if consultants were truly independent, the commissioner in model C would still be asking the questions and could still choose to overlook specific questions.

An organisation's choice of A over B – i.e., whether to conduct in-house or contract out – is likely to depend on the evaluation function's size and resources, where A is likely to be preferable when feasible, as it may be more efficient when the same unit both designs and carries out evaluations.

---

[8] External consultants may, for professional and ethical reasons, not engage in this behaviour.

This also reduces transaction costs arising from procurement. The challenge with model A, however, is that it is difficult to have all the relevant staff in-house for all the different topics that the function needs to cover. This could perhaps explain why many bilateral evaluation functions rely on external consultants, as they cover many evaluation topics.

However, this is not to say that evaluations commissioned by a stakeholder (i.e., model C or D) are not without merit.

Most development agencies have a plethora of evaluations that are commissioned for instrumental or conceptual uses by programming units or grant managers. These evaluations can serve both accountability and learning purposes. For example, a development agency may commission an evaluation of a partner project (accountability purposes) or an evaluation of its own projects (learning purposes). These types of evaluations often outnumber the evaluations commissioned by an independent evaluation function.

The relative importance of independence (and the perception of independence), as it links to credibility and thus use, is affected by the evaluation purpose and intended use (and users). When improvement and learning are the focus, the involvement of the targeted users may enhance credibility more than "independence". If the main user of the evaluation is the public, or other accountability bodies, perceived independence will be of utmost importance – particularly in contexts of political debate and aid scepticism. The role of independence in influencing credibility is thus mediated by both the purpose of the evaluation and its context.

Finally, there is an increasing focus on the involvement of impacted populations in evaluations, and it is clear that the power dynamics of "traditional" donor/recipient-type relationships create and reinforce certain types of bias. Such biases and blind spots in programming can be replicated in evaluation, or even reinforced. The notion of "independence" therefore needs to be viewed from a broader perspective where it is in service to credibility. In many contexts, participatory evaluations – involving people who are meant to benefit from an intervention, staff involved in the implementation, or others – may provide a more credible and less biased outcome.

7. Concluding remarks

Independence is both a concept and a practical element of evaluation quality. It should be understood as a means to an end and in relation to both perceived and real biases, as well as to overall credibility. This note has laid out some framing of the issues to encourage an unpacking of quality elements and a nuancing of how best to use the tool of independence – applying it differently to ensure credibility in different contexts, both in terms of avoiding implicit and explicit bias as much as possible, and to address perceived biases that impact both credibility and use.

References

- Department for Evaluation. (2014). *Can we demonstrate the difference that Norwegian aid makes?* Norad.

- Easterly, W. (2002). The cartel of good intentions: The problem of bureaucracy in foreign aid. *The Journal of Policy Reform*, 5(4), 223–250.

- Kupiec, T., Celińska-Janowicz, D., & Pattyn, V. (2023). Understanding evaluation use from an organisational perspective: A review of the literature and a research agenda. *Evaluation*, 29(3), 338–355.

- Lindkvist, I., & Dixon, V. (2014). To 'feel good', or to 'do good'? Why we need institutional changes to ensure a results focus in Norwegian development assistance. *Journal of Development Effectiveness,* 6(4), 350–360.

- Michaelowa, K., & Borrmann, A. (2006). Evaluation bias and incentive structures in bi-and multilateral aid agencies. *Review of Development Economics*, 10(2), 313–329.

- Niskanen, W. A. (1968). The peculiar economics of bureaucracy. *The American Economic Review*, 58(2), 293–305.

- OECD. (2010). *Quality standards for development evaluation* [DAC Guidelines and Reference Series]. OECD Publishing. https://doi.org/10.1787/9789264083905-en.

- OECD. (2023). *Evaluation systems in development co-operation 2023*. OECD Publishing.

- OECD/DAC. (1991). *Principles for evaluation of development assistance* (Principles No. 91/208).

- Pritchett, L. (2002). It pays to be ignorant: A simple political economy of rigorous program evaluation. *The Journal of Policy Reform*, 5(4), 251–269.

- Reinertsen, H., Bjørkdahl, K., & McNeill, D. (2022). Accountability versus learning in aid evaluation: A practice-oriented exploration of persistent dilemmas. *Evaluation*, 28(3), 356–378.

- Svensson, J., & Bjorkman, M. (2007). *Power to the people: Evidence from a randomized field experiment of a community-based monitoring project in Uganda*. The World Bank.

- Vedung, E. (2012). Six models of evaluation. In *Routledge handbook of public policy* (pp. 387–400). Routledge.

Department for Evaluation

norad.no/evaluation