3. REFLECTIONS ON INDEPENDENCE IN EVALUATION

HANS-MARTIN BOEHMER, Senior Manager
XIMENA FERNANDEZ ORDONEZ, Evaluation Officer
NEHA SHARMA, Consultant

Independent Evaluation Group
The World Bank Group

This paper integrates a synthesis and analysis of the discussions held on the online platform leading up to and during the Third International Conference on National Evaluation Capacities.[27] It is also based on an initial paper that was produced to serve as background for the presentation on independence and related discussions during. The background document and the discussions focused on the relationship between independence and evaluation, seeking to contribute to linking how credibility and use interact with independence. The background paper was based on an analysis of different examples of how organizations and countries deal with the issue of independence in their M&E systems.[28]

INTRODUCTION

Countries, development organizations and scholars often struggle to determine what independence in development evaluation really implies, what the optimal level of independence for an evaluation function should be and how to achieve that level. In these conversations, the independence of an evaluation function is often discussed in isolation from the ultimate role of evaluation. Also, independence is often portrayed in absolute terms. That is, evaluation functions are either independent, or they are not.

27 The paper does not reflect the opinions of the Independent Evaluation Group of the World Bank. It also does not necessarily reflect the views of its authors, since it attempts to summarize and analyse the conversations held during the Third International Conference on National Evaluation Capacities.

28 Analysis of Country M&E System cases was based on IEG Evaluation Capacity Development Working Paper Series; see ieg.worldbankgroup.org/evaluations/evaluation-capacity-development-ecd.

The conversations before and during the conference highlighted the view that in the development realm, independence in evaluation is only relevant for its help in supporting evaluation's ultimate function: improving development results. Independence is not an end in itself. Therefore, different levels and types of independence might be appropriate for different situations. There is, unfortunately, no recipe or blueprint for achieving 'optimal independence'.

This paper also posits that independence should be seen as a characteristic that helps reduce the biases that an evaluation function might have (or be perceived as having). Reducing such biases should increase an evaluation's credibility, which in turn should increase the use of its evidence to feed decisions. The assumption is that 'evidence-based decision-making' will ultimately improve development results.

Independence is relevant to evaluation because assessing the results of development projects, programmes and policies is complex, and many biases can emerge in the evaluation process. However, the relevance of each of these biases—and therefore the need to deal with it in a particular situation—depends on whether or not it affects relevant actors' perceptions regarding the credibility of the evaluation function and the evidence it produces. Evaluation functions face many biases—some real and some perceived. Independence cannot preclude some of them; they just need to be managed. And even though independence is an important component that influences credibility, merely establishing independence may not be enough. Similarly, credibility is not the sole determinant that leads to the use of evidence. In fact, independence in evaluation is often seen, not necessarily correctly, as endangering the adoption of the evidence produced by evaluations.

The experiences of many countries' and development organizations' efforts to influence development results through evaluation shows that a battery of responses, often articulated through institutional M&E systems, is helpful in aligning independence with other important characteristics that support evidence-based decisions. The Independent Evaluation Group of the World Bank regards independence as one of the keys to ensuring evaluation credibility and the use of the evidence it produces. Other organizations and countries that have different realities have chosen different paths; M&E systems can help find an 'optimal' or 'useful' type and level of independence for a particular situation.

The first section of this paper explores why independence is important in evaluation. The second section examines the definition of independence and its relevant dimensions. The third section focuses on the current debates about independence, including the trade-offs, overlaps and supportive functions between independence and commonly implemented solutions to uphold it.

WHY IS INDEPENDENCE IMPORTANT IN EVALUATION? THE LINK BETWEEN INDEPENDENCE AND RESULTS

> *"Independence is not an end in itself (but it is meant to ensure) "objective, impartial, uncompromized and unbiased findings that will make the evaluations credible and valuable."*
>
> ROBERT PICCIOTTO (2013)

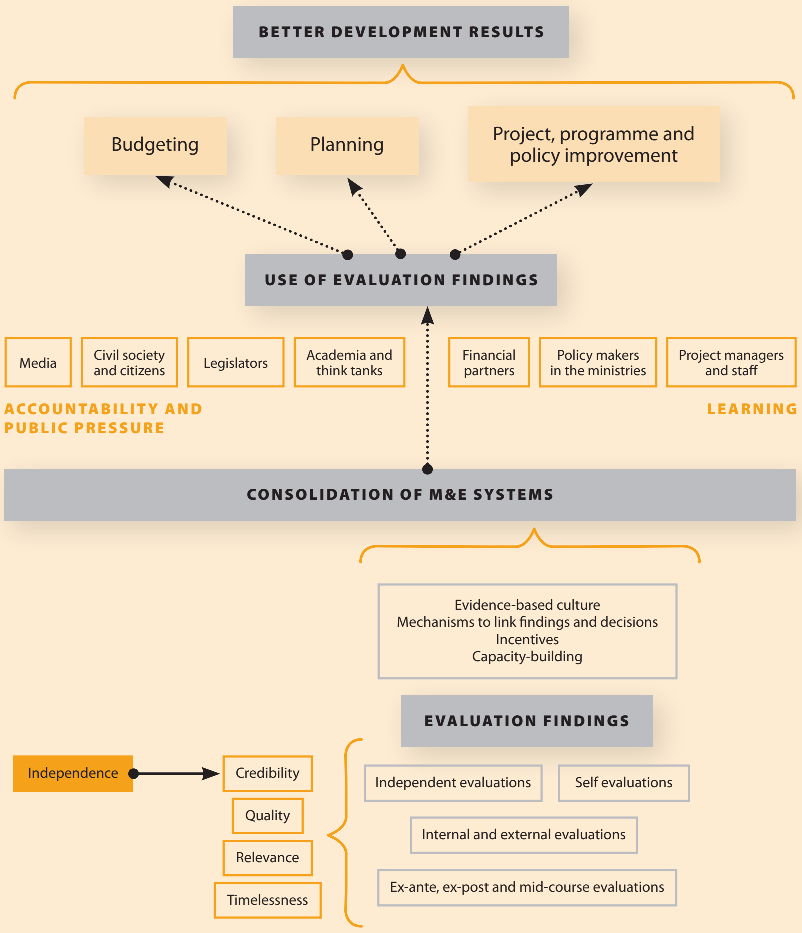
"The idea of independence has a long association with evaluation (and in many cases is seen as) an essential element of good evaluation practice" (INTEVAL 2011, 1). Independence becomes particularly relevant in development evaluation because of the complex nature of development policies, programmes and projects; the poor availability of data; the multiplicity of actors; and the perceptions that particular interests might have in confirming positive results rather than truly assessing what happened and why. This evaluative context can differ among areas; for example, in medicine, data is more widely available and there is greater opportunity to conduct well-controlled experiments than elsewhere.

Discussions of independence in evaluation often leave out the ultimate role that evaluation itself is meant to play. Evaluation findings can help countries and agencies know what progress has been made (if any), why it did or did not happen, the mechanisms that aided it and, in some cases, what portion of those results can be attributed to specific project, programme or policy feature (Thomas and Tominaga 2010, 376). Evaluations can supply information for feedback loops and nudge real-time improvements by incorporating past learning and lessons into later planning. Evaluation can help improve the design and implementation of ongoing projects, programmes and policies; promote the scale-up and replication of successful practices; avoid mistakes; and incorporate learning into higher-level planning, budgeting (or prioritization) and design of programmes. Ideally, evaluations should enable stakeholders to ask tough questions, challenge assumptions, allow for changes in courses of action, learn progressively, make space for reforms and inform policy debates (Boehmer, in Rist 2011).

What purpose does independence serve in development evaluation and how does it operate? If we were to map a theoretical 'results chain' for independent evaluations, we would see that independence is meant to prevent bias (promote the objectivity) of an evaluation function, which in turn strengthens its credibility (see Figure 1). This increased credibility should result in a rise in the utilization of the evidence. Better-informed decisions or evidence-based decision-making should then improve development results. In some cases, the evidence influences decision makers directly. In others, it mobilizes accountability mechanisms that influence them indirectly.

Depending on the actors that the evaluation function is trying to influence, not all biases will hold the same importance. The biases that will matter will be those that affect the credibility of the evidence in the eyes of the actor who needs to act on it. Independence, however, is not the only factor that affects credibility, and credibility might not be enough to ensure the actor incorporates an evaluation's findings into their decision-making process. Lastly, though evidence-based decision-making could improve development results, it is likely not the only determinant.

FIGURE 1: A THEORETICAL RESULTS CHAIN FOR INDEPENDENT EVALUATION



In his 2013 study of evaluation independence, Picciotto found that "without independence, evaluation is perceived by the public to be subservient to vested interests." When evaluation results appear to be tied to personal gain, profits or corporate interests, the ability to achieve desired results becomes seriously compromised. Conversely, establishing independence can validate results, which may significantly improve the ability to achieve desired results (Picciotto 2013, 22).

Because of this connection, the discussion on independence in evaluation cannot be separated from the discussion on avoiding bias. Independence can decrease conflicts of interest, where the evaluator is "'in bed with' the program being evaluated [which is] typical of much program monitoring by agencies and foundations where the monitor is usually the godfather of the program, sometimes its inventor, and nearly always its advocate at the agency" (Scriven 1991, 192-193). Independence—or true freedom from political, financial or personal objectives—is supposed to render unbiased findings. Both real and perceived biases are relevant, and so both perceived and real independence is, too.

The discussion on independence also needs to be linked to the actors that the evaluation is trying to influence and their perceptions regarding the biases that affect credibility. By increasing the credibility of evidence, key stakeholders (e.g. parliaments, opposition parties, civil society organizations, citizens, universities, think thanks and the media) have better tools available and might be more likely to hold governments and agencies accountable and exert public pressure for course corrections (Mackay 2007). Further, managers and planners can see more clearly their mistakes and missed opportunities and pre-emptively incorporate such information into their decisions.

Biases and other constraints can emerge at every stage of an evaluation, including when the subject is chosen, the questions determined, the methodology designed, the engagement and consultation strategy with stakeholders set, the information analysed, the evaluator chosen, the evaluation terms of reference set, the evaluation implemented, the recommendations constructed or the results are reported and presented. Factors such as who is paying for the evaluation, how the evaluation is managed and who the evaluators are, are also critical regarding the emergence of biases.

Independence can abolish or reduce biases and increase objectivity. However, independence—or at least the ways in which it is often understood—also has the potential to diminish the expected effects of evaluation on results.

Independence is not the only way to improve credibility. For example, making the data used for evaluation available (i.e. giving individuals, academia, think tanks and others the chance to analyse it) is one way to increase the legitimacy of the evidence, or at least diminish questions regarding its objectivity. When external third parties have the opportunity to use their own methods and draw their own conclusions and recommendations, the actual evaluation findings can be validated, refuted or understood as one interpretation.

Moreover, independence is a necessary but not a sufficient condition for evidence to affect results. Evaluation findings can only play a predominant role in promoting effective and efficient projects, programmes and policies if they are incorporated into the relevant decisions around them. Contrary to the assumption that credible evidence will influence decisions, many aspects beyond credibility affect use. Key factors in determining the role that evaluations play in achieving better results include the relevance of the topics evaluated, the evaluation's guiding questions, the timeliness and quality of the evaluation, the dissemination of the findings, the existing structures link findings and decision-making, the incentives in place, the level of capacities to engage with evidence and the culture of an organization.

In conclusion, though independence could help reduce many biases that affect credibility, other biases that are not possible to avoid through independence will need to be managed; independence is only one of the ways to increase the chances of the evidence from evaluations influencing results. A country or an organization first needs to understand the perceived or real biases evaluations face and then choose the type and level of independence that appropriately responds to a given situation.

WHAT IS INDEPENDENCE? FORMAL DEFINITIONS AND DIMENSIONS

> *Independence: Freedom from external control or influence.*
>
> NEW OXFORD AMERICAN DICTIONARY

Though there is a lot of discussion on independence in evaluation, there is rarely agreement on what independence really is. The definition of the United Nations Evaluation Group (UNEG) regards an evaluation function as independent if it is "objective, free from undue influence, and has full authority to submit reports directly to appropriate levels of decision-making" (UNEG 2005). Implied in this definition is the directive that "management must not impose restrictions on the scope, content, comments and/or recommendations of evaluation reports." To prevent conflicts of interest, UNEG also asserts that evaluators "must not be directly involved in policy-setting, design, implementation or management of the subject of the evaluation" and adds that "evaluators should have full access to all relevant information required for the evaluation". Somewhat less straightforward, but more extensive, is the definition of the Organisation for Economic Co-operation and Development's Development Assistance Committee's (OECD-DAC), which considers an evaluation independent if it is "carried out by entities and persons free [from] control of those responsible for the design and implementation of the development intervention" (OECD-DAC 2002). Similar to UNEG, the OECD-DAC definition includes 'full access to information' and 'full autonomy' as necessary characteristics of an evaluation's design, implementation and reporting for it to be considered 'independent'.

Both structural and individual independence (also called functional or intellectual) are relevant in evaluation (Heider 2014 and Picciotto 2013). Even though this paper focuses on the institutional and organizational dimensions of the relationship between independence and evaluation, it should be noted that for an evaluation to be considered truly independent, these requirements must apply to the evaluation function or unit and to the actual individual or team conducting an evaluation.

Individual evaluators also suffer from numerous natural biases and risks that can corrupt findings, such as anchoring, framing and attention bias. Anchoring refers to focusing on the one aspect of a programme that confirms preconceived notions about the project; framing refers to drawing different conclusions from the same information that others would have available, depending on how or by whom that information is presented; and attention bias refers to paying attention to emotional stimuli instead of facts. Ignorance, ideology and inertia can also be important biases affecting both organizations and individuals (Boehmer, in Rist 2011). An evaluator needs 'independence of mind', the state of mind that permits one to provide an opinion without being affected by influences that compromise professional judgment, thereby allowing an individual to act with integrity and exercise objectivity and professional scepticism (International Federation of Accountants 2003).

The question that arises is how countries and organizations should use these definitions and dimensions when structuring evaluation functions. Independence is still often seen in absolute terms: evaluation functions are seen as either independent or not. In reality, complete independence, or the avoidance of all biases, is hard to achieve. So should countries and organizations aspire to 'tick' all these boxes? Current discussions around this subject have taken quite a nuanced approach, increasingly linking independence to organizational learning theories and specific contexts (e.g. relevant decision makers, perceived and real biases) (Picciotto 2013). Countries and development organizations have also taken very different routes to ensuring the independence, credibility and use of evidence.

Despite the availability of good definitions, independence is still often oversimplified and equated or opposed to other terms and constructs. The next section of this paper explores some of the common views regarding evaluation independence, how to reduce biases and the likely effects these can have on credibility, use and results. Some of these 'solutions' might actually go against the use of evidence.

THE CASE OF THE INDEPENDENT EVALUATION GROUP: INDEPENDENCE WITHIN

In the World Bank Group, structural (organizational) independence (actual and perceived) is seen as one of its two pillars for the credibility of its evaluation function and its influence on the World Bank Group's results (the other pillar is the 'quality and relevance' of its work). Independence contributes to the overall governance of the World Bank Group by allowing stakeholders to use unbiased and reliable findings. According to its 2011 self-evaluation, the Independent Evaluation Group's (IEG) organizational systems, reporting structures and procedures are consistent with those established by the Evaluation Cooperation Group Good Practice Standards for independence.

But IEG is not truly separated from the World Bank; it enjoys a kind of 'independence within'. It is an independent evaluation function that remains inside the World Bank architecture. This independent evaluation function strongly complements self-evaluation at the World Bank. In this position, IEG—much like other evaluation departments—faces internal pressures to function as an internal self-evaluation group instead of an independent group. In this situation, issues such as engagement become critical. However, in 2004, an external review noted that engagement with management did not undermine independence; "on the contrary, such interaction should be increased to ensure the usefulness of evaluation products." (Continued on p. 60)

The Case of the Independent Evaluation Group: Independence Within (continued from p. 59)

IEG INDEPENDENCE CRITERIA	
CRITERION	**FACTORS THAT HELP MEET CRITERION**
Organizational Independence	IEG reports directly to the World Bank Board through the Committee on Development Effectiveness (CODE) and is thus organizationally independent from management and operational staff whose activities are being evaluated. IEG's scope of responsibility extends, without restriction, to all the determinants of the World Bank Group's operational results.
Behavioural Independence	IEG's Work Program and Budget are endorsed by CODE and approved by the Board. The IEG budget is separate from management budgets, and in the end, management does not have authority over IEG's budget or its use. IEG's reports are transmitted to the Board through the Director General for Evaluation, without any clearance from management. Although management is given the opportunity to review and comment on draft reports, IEG decides how and whether or not to address such comments.
Avoidance of Conflict of Interest	IEG staff does not evaluate activities that they were previously responsible for or were involved in. The head of IEG and its Director Generals are not eligible for employment in other positions in the World Bank Group or for consulting assignments.
Protection from Outside Influence	IEG's three-year rolling consolidated work programme and budget are prepared independently of management for endorsement by CODE and approval by the Board.

The IEG has also attempted to increase the influence of its evidence on World Bank decisions—and ultimately its effect on development results—through the Management Action Record (MAR), a system that mandates and facilitates incorporating emerging recommendations into the Bank's policies and programmes, and in a sense, mediates the relationship between IEG and management. MAR allows IEG to track the progress made by World Bank Group management in adopting its recommendations from sector, thematic and corporate evaluations.

MAR has helped both IEG and management by: improving the quality of IEG recommendations (by providing clearer links to findings, prioritizing, improving clarity and specificity, and integrating considerations of feasibility and cost-effectiveness); strengthening engagement and building understanding and agreement with World Bank Group management while drafting recommendations; increasing the number of specific actions to implement IEG recommendations and clarifying timelines and monitoring arrangements; enhancing assessment of progress on implementation; and reducing inconsistencies between IEG and management ratings.

IEG and World Bank Group management jointly began to reform MAR in 2011. After intensive coordination and testing with the three World Bank Group institutions, the new standardized MAR system for tracking recommendations across institutions was rolled out in April 2013. Currently, MAR provides stakeholders with a central repository of findings, recommendations, management responses, detailed action plans and implementations. IEG will continue to provide access and training to all interested members of the Board and World Bank Group management and will also develop a series of briefings on the available tools. As part of IEG's commitment to enhance transparency and access for external stakeholders, IEG has made MAR data available on its external website. The MAR system currently houses 193 individual IEG recommendations across the World Bank Group (World Bank, International Finance Corporation and Multilateral Investment Guarantee Agency). In 2013, IEG followed up on 98 recommendations active in the system.

These same criteria are shared by the Evaluation Cooperation Group, which was founded in 1995 by the heads of evaluation departments of multilateral development banks. The members of the Evaluation Cooperation Group include the evaluation departments of: African Development Bank, Asian Development Bank, European Bank for Reconstruction and Development, European Investment Bank, Inter-American Development Bank Group, International Monetary Fund and World Bank Group.

WHO ARE THE FRIENDS AND FOES OF INDEPENDENCE? PERCEPTIONS, REALITIES AND MYTHS

> *"Independence combined with disengagement increases information asymmetry, ruptures contacts with decision makers and restricts access to relevant sources of information … thus, the basic challenge of evaluation governance design consists in sustaining full independence without incurring isolation."*
>
> ROBERT PICCIOTTO (2008)

There are numerous preconceptions regarding the relationship between independence and other principles, concepts or constructs. 'Engagement' is often seen as opposed to independence, and 'separation' and 'distance' are often portrayed as its ally. Independence is also often seen as a control function, linked solely to accountability. And to preserve independence, *ex post*, externally conducted evaluations are often the chosen tool. Independent evaluations are readily used as an indication of accuracy and of good quality.

These constructs are linked to attempts to reduce biases and increase organizational and behavioural independence. But sometimes, actions taken in the name of independence can actually have a detrimental effect on the objectives that independence is pursuing. Also, some of these concepts and constructs contain some truth, some do not, and the complexities of reality often preclude such binary determinations. Some of these concepts overlap, some contradict, and others reinforce the notion of 'independence'. This paper does not attempt to resolve such tensions, but rather endeavours to shed some light on how they affect the credibility and use of evidence.

Independence is often understood as 'separation' or lack of engagement

Independence is often interpreted as drawing and maintaining a clear line between the evaluation function or evaluator, the topic of evaluation and those responsible for the design and implementation of the policy, programme or intervention under evaluation. It is often assumed that maintaining separation or 'an arm's length' from decision-making can help prevent bias and provide objectivity to the evaluator. The fear is that intentional or unintentional biases might develop if a relationship develops between the evaluation unit and the managers of the programmes or interventions being evaluated. However, the line between the evaluator and the programme or project is often blurry.

The Evaluation Cooperation Group recognizes that independence does not mean isolation: cross-fertilization of knowledge and experience enriches operational and evaluation activities. Experts reinforce the idea that "evaluation cannot fulfil its potential unless it connects effectively to its management, the supreme authorities that govern the organization and the broader society" (Picciotto 2013).[29] Additionally, "selecting the right topics and providing timely inputs require deep understanding of the institution that can be gained through close engagement." Open and transparent conversations among evaluators,

29 "But in doing so it should maintain its objectivity, exercise full freedom of inquiry and resist capture. Evaluation needs to remain functionally and structurally independent" (p. 22).

programme managers, implementers, beneficiaries and civil society regarding all aspects—from programme design through to discussion of programme outcomes and findings—may improve an evaluators' credibility and enhance the evaluation's learning and problem-solving functions.

Establishing an effective balance between independence and engagement could indeed help link evaluations to results. Several cases have shown that although it might take several iterations, it is possible to structure engagements in such a way that they do not violate independence. "Engagement … provides internal credibility and can lead to constructive change" (World Bank 2013).

Evaluations are also often considered independent when the evaluation function or evaluator sits outside of the institution or is not part of its organizational structure. Being 'far' could offer objectivity, which is why many organizations prefer outside/external evaluators. However, though outsiders can provide a fresh perspective, distance alone does not ensure independence. Simply existing external to an organization does not ensure that there are no political, personal or financial incentives that would invalidate independence. Further, although external evaluators can bring third-party insights, they still face biases. For example, fee dependency of external evaluators can constrain independence as much as proximity; the incentive set-up may not always be conducive to independence (Picciotto 2013). External consulting firms and evaluators with a strong client orientation and business sense often have powerful incentives to earn contract renewals and obtain new contracts, which can corrupt independence by prioritizing customer satisfaction over analytical rigour.

Another concern with external evaluators is that their understanding of the programme or its context may not be sufficient enough to provide useful findings. Additionally, external evaluators may have fewer incentives than those conducting a self-evaluation to go the extra mile to make the evaluation findings useful for learning. Yet another constraint on the use of evidence is the fact that external evaluators do not always have a firm grasp of implementation realities and may establish questions, frameworks or methodologies that are not realistic or relevant.

In the past, independent evaluation was often associated only with accountability and interpreted as an external control

Independent evaluation functions are often considered to be a policing or 'watchdog' function, and are therefore only linked to accountability. This view is reinforced by the fact that evaluations often use language that can hinder communication of the main messages. It is important that independent evaluators present balanced reports, highlighting not only areas that need improvement, but also the constraints and other implementation roadblocks that organizations face. Balanced, realistic reports will increase organizations' receptivity to findings and to evaluations in general.

'Learning' is increasingly being recognized as an integral objective of evaluation. "In evaluation we do not aim to 'learn in the abstract' or test theoretical concepts, but learn about what happened during the process of implementing a policy, strategy, program [sic], or project… [Learning and accountability] are … two sides of the same coin: accounting for what happened during implementation when working towards achieving the plans set out in a project design—an agreement and commitment between at least two parties—is the source for learning about the replication of success and avoidance of unnecessary mistakes on that journey to realize expected outcomes" (Heider 2014). Independence does not mean that the work is done when findings are published. Influencing results means that evaluation functions include the responsibility to actually try to foster the adoption of their findings and recommendations.

Independence is also often conflated with an evaluation's type and timing

Although less discussed, there may be presumptions that additional biases emerge if the evaluation function comes in early in a project or programme's life cycle. This perception is related to the question of whether it is possible to evaluate work that was influenced by the evaluation function itself. However, as evaluation is an integral part of the policy cycle and tries to influence decision-making (even if only through learning from other evaluations), this problem is present every time findings affect a decision, even indirectly.

Although commissioning evaluations after programme implementation may help reduce some bias and conflicts of interest, it does not guarantee their avoidance. And an evaluation is not necessarily independent simply because it is conducted *ex post*. Similarly, the opposite might not mean complete lack of independence. An evaluation function can conduct independent formative and mid-course evaluations. This would allow for incorporating corrections based on the findings. In other words, it would feed into the feedback loops for programme learning.

If evaluation functions become the history department of a programme, they may not have a significant impact on results: inputs will come too late to allow the organization to change course, and the context may change too fast for lessons to still be relevant. If evaluations do not link appropriately to feedback loops, there is a serious risk that such evaluations may compromise development results and lead to "unjustified replication of popular fads based on anecdotal evidence" or prevent "structured learning … [or] replication of innovations that may be performing better than is appreciated" (World Bank 2013). Other types of engagements, such as assessments of evaluability practiced by the Inter-American Development Bank, could help develop more accurate programme or policy results frameworks, indicators[30] or M&E plans, leading to programme design improvements. Though there are risks and biases related to an earlier engagement that need to be managed carefully, in certain situations and contexts the potential benefits to results may outweigh these risks.

Self-evaluations are often portrayed as the enemy of independence

Self-evaluations and independent evaluations often have complementary functions and can support each other. Conducted by the implementing organization to evaluate its own work, self-evaluations benefit from deep knowledge of the organization, its mission and its

30 Poor quality at entry is a strong predictor of poor eventual outcome, although there might be other factors correlated with that relationship.

programmes. Self-evaluations are more likely to overcome information asymmetries than independent ones (DFID 2008). They are also more likely to be well-received by management and programme implementers. Therefore, self-evaluations' findings have a higher probability of being used for mid-course corrections. Self-evaluations are also important elements of an organization's culture of seeking and using evidence (Mayne 2008 and INTEVAL 2011).

However, there are valid concerns regarding an organization's ability to conduct self-evaluations that are free of biases, given managerial relationships, ownership of or accountability towards projects being evaluated, and the link between results and funding requirements. Moreover, there are moral hazards of judging one's own work that are related to a reluctance to accept failure. It is crucial to complement and balance self-evaluation with checks and balances or an independent evaluation function tasked with attesting to the validity of self-evaluation findings (Picciotto, NEC 2013 community of practice).

There is no direct link between independence of an evaluation and its quality or accuracy

Without independence, the quality of an evaluation will usually be questioned. However, "independence on its own does not guarantee evaluation quality: relevant skills, sound methods, adequate resources and transparency are also required. Independent but incompetent evaluations can be misleading, disruptive and costly" (Picciotto 2013, 20). On the other hand, self-evaluations can be of very high quality. In any evaluation, it is critical to ensure that the evaluators have the appropriate skills and the technical, contextual and procedural knowledge to conduct the evaluation. The evaluator selection process and the quality control mechanisms for evaluations of any type (e.g. steering or technical groups) are key in this regard. Independent evaluations are often also perceived as being more accurate than self-evaluations.[31] However, other types of biases not related to proximity to the programme may be equally present and skew the evidence they produce.

The proposed responses to decreasing biases or preserving independence can go against the core of what independence tries to achieve. In some cases, the risks of emerging biases can be managed. Under certain circumstances, sacrificing independence for better results may be a reasonable option. At other times, even if the biases are merely perceived, the danger is just too big. Real or perceived, these tensions and trade-offs are usually better managed within an M&E system rather than individually.[32]

CONCLUSIONS AND QUESTIONS GOING FORWARD

Independence has traditionally been a very important aspect of development evaluation. Integrally linked to the credibility and use of evidence, the relevance of independence cannot be detached from the ultimate goal of evaluation: to improve development results.

31 As Ken Chomitz, Senior Adviser at IEG explains, just like in econometrics, the "unbiased estimator can be the one with the biggest error."

32 The analysis of the Country M&E System cases was based on the Independent Evaluation Group ECD Working Paper Series, available at ieg.worldbankgroup.org/evaluations/evaluation-capacity-development-ecd.

However, some of the ways in which independence is interpreted or operationalized can have neutral or negative effects on credibility and use, and therefore do not assist with improving results. There has traditionally been a false dichotomy between independence on the one hand and learning and incorporating evaluative evidence for decision-making on the other (Picciotto, NEC 2013 community of practice). But if independence is to have an effect on results, evaluation functions need to have an important learning role and not operate just as a control function.

The discussions and cases analysed as the basis for this article highlighted that, when it comes to independence, one size does not fit all. Case studies and examples can provide inspiration and illustrate different approaches organizations and countries have taken with regards to independence. However, there is no single model or blueprint to attaining independence, nor is there a definitively appropriate response to particular biases or threats to independence. However, some lessons emerge:

Engagement with programmes and managers does not necessarily endanger independence; it can actually help with the credibility and use of evidence. Conversely, separation and distance between management and evaluation can endanger the adoption of evidence. Irrespective of an evaluation's independence, timing and type, its quality should always be a concern because it is directly related to credibility. Often, the most successful evaluation systems combine self-evaluations and independent evaluation functions. Risks and biases need to be managed carefully, but this is true for all aspects of evaluation, not just when dealing with increased engagement or self-evaluation.

M&E systems can play a significant part in better linking and balancing independence, the use of evidence in feedback loops and learning from lessons for future projects, programmes and policies. Independence is a very important tool that countries and development organizations have in fostering credibility and use, but it cannot be the only one. Designing M&E systems often entails making tough decisions, including the type and level of independence. In making those choices, countries and organizations need to ensure that they are not endangering adoption and use in the name of independence. M&E systems tend to respond to the reasons they emerged (decision-making needs, relevant actors, levels of capacity and readiness for evaluation), and they can play a critical role in fostering evaluations that are credible, relevant, timely and of good quality. In order to design an M&E system that is a good fit for its context, it is critical to incorporate diagnostics that facilitate a better understanding of the reality on the ground.

For some systems (e.g. the Independent Evaluation Group, the United States, and the United Kingdom's Department for International Development), independence is a pillar. For others (e.g. Canada), independence is sought after[33] but is not the main feature. Even among those systems that have chosen to prioritize independence, many have started transitioning towards designing more straightforward links to foster the incorporation of evidence

33 The analysis of the Country M&E System cases was based on the Independent Evaluation Group ECD Working Paper Series. See the cases at ieg.worldbankgroup.org/evaluations/evaluation-capacity-development-ecd.

into decision-making. The links established among the evaluation findings, project and programme improvement, and functions such as planning and budgeting, are critical. Many independent evaluation functions, such as those of the National Council for the Evaluation of the Social Development Policy in Mexico and the Independent Evaluation Group, have started to set up robust mechanisms to follow up on the recommendations of evaluations.

Different institutional arrangements, processes, incentive structures, M&E tools and methodologies have helped address the many risks and biases related to credibility that were particularly relevant for their contexts. Variations in cultures, time periods and topics need different degrees of independence. As M&E systems evolve, establishing better links between independence and results still presents an opportunity for innovation and advancement in the M&E field.

One of the most interesting questions going forward is how other countries and organizations have achieved an optimal level of independence that allows for improving development results. Which arrangements have been successful in linking independence, credibility, and the use of evidence? What competencies and incentives have been successful in fostering these links? We look forward to the further debates on this topic.

REFERENCES

Heider, Caroline, "Independence for Accountability and Learning", Forthcoming.

International Federation of Accountants. 2003. *Handbook of International Auditing, Assurance and Ethics Pronouncements*. International Federation of Accountants.

Mackay, Keith Robin. 2007. *How to Build M&E Systems To Support Better Government*. Independent Evaluation Group. Washington: World Bank.

Mark, Katharine, and John R. Pfeiffer. 2011. "Monitoring and Evaluation in the United States Government: An Overview". ECD Working Paper 26. Washington: World Bank.

Mayne, John. 2008. *Building an Evaluative Culture for Effective Evaluation and Results Management*. Institutional Learning and Change Initiative (ILAC).

Organisation for Economic Co-operation ans Development, Development Assistance Committee (OECD-DAC), Working Group on Aid Evaluation. 2002. "Glossary of Key Terms in Evaluation and Results Based Management". Available at oecd.org/development/peer-reviews/2754804.pdf. Last accessed 5 February 2014.

Picciotto, Robert, and Eduardo Wiesner Durán (Editors). 1998. *Evaluation and Development: The Institutional Dimension*. Vol. 1. Transaction Publishers.

Picciotto, Robert. 2008. "Evaluation Independence at DFID". Independent Advisory Committee for Development Impact.

Picciotto, Robert. 2013. "Evaluation Independence in Organizations". *Journal of MultiDisciplinary Evaluation* 9 (20):18-32.

Rist, Ray (Editor). 2011. *A Vision for Development: Dialogues on the Work of Vinod Thomas*. Independent Evaluation Group. Washington: World Bank.

Scriven, Michael. 1991. *Evaluation Thesaurus*. Sage Publications.

Thomas, Vinod, and Jiro Tominaga. 2010. "Evaluation for Better Development Results." *Journal of Development Effectiveness* 2(3):371-386.

United Nations Evaluation Group (UNEG). 2005. "Standards for Evaluation in the UN System". Available at uneval.org/papersandpubs/documentdetail.jsp?doc_id=22. Last accessed 5 February 2014.

World Bank Independent Evaluation Group. 2011. "Self-Evaluation of the Independent Evaluation Group". Washington: World Bank.

World Bank Independent Evaluation Group. 2013. "Management Action Record Database." Washington: World Bank.