

Conceptual Framework for Developing Evaluation Capacities Building on Good Practice in Evaluation and Capacity Development¹

Caroline Heider

Introduction

The purpose of this chapter is to contribute to a better understanding of the complexities of evaluation capacity development and suggest ways of approaching them, within an institutional context or at national level. It applies “good practice” in capacity development to attain “good practice” in evaluation so that it can serve its purposes of accountability and learning. Evaluation can and should help governments and organizations achieve their objectives. This paper builds on established “good practice” in these two professions and on the experience of evaluating capacity development.

Both professions – that of evaluation and of capacity development¹ – are multifaceted and evolving, which means this chapter captures only a segment of the whole at a certain point in time. Capacity development practitioners continue to develop and refine concepts and practices; the 2006 publication *The Challenge of Capacity Development*² is subtitled “working towards good practice” in full recognition of the challenges that are still ahead. In a similar way, the evaluation profession continues to grow.

This chapter comes at an important point in time. The demand for evaluation is high, but increasingly critical of evaluation practices that are not demonstrating independence, credibility and utility. In addition, following the Paris Declaration and the Accra Agenda for Action, there is an increasing demand for evaluation capacities in partner countries. For instance, the evaluations of the “Delivering as One” initiative of the United Nations will be country-led. Therefore, developing evaluation capacities is important for a range of stakeholders in the development and humanitarian communities. This chapter aims to provide these stakeholders with a conceptual framework to support their efforts.

The chapter is divided into three main sections: one to discuss evaluation capacities at three levels, which overlays the evaluation principles with capacity development ideas, one to discuss the capabilities and processes to instil them, and finally one to discuss examples and draw out suggestions for stakeholders working on evaluation capacity development.

Evaluation Capacities at Three Levels

Evaluation Principles

The evaluation principles that are used here date back to those defined at a meeting of evaluation professionals in 1991³ and probably to times before then. The terms are defined in the annex to this chapter. They have become integral to the evaluation profession at large.⁴ For instance, independence has taken central stage in the multilateral development banks when the evaluation function at the World Bank started reporting to the executive board rather than the president; a change that was replicated in regional development banks and that stimulated discussions in, among others, the United Nations. In 2008, the framework for peer reviews of evaluation functions in the UN system⁵ raised three principles to be central to evaluation: independence, credibility and utility.

¹ This paper was prepared as an input to the forthcoming publication by the International Development Evaluation Association (IDEAS) and the World Bank’s Independent Evaluation Group (IEG), to be released in 2010.

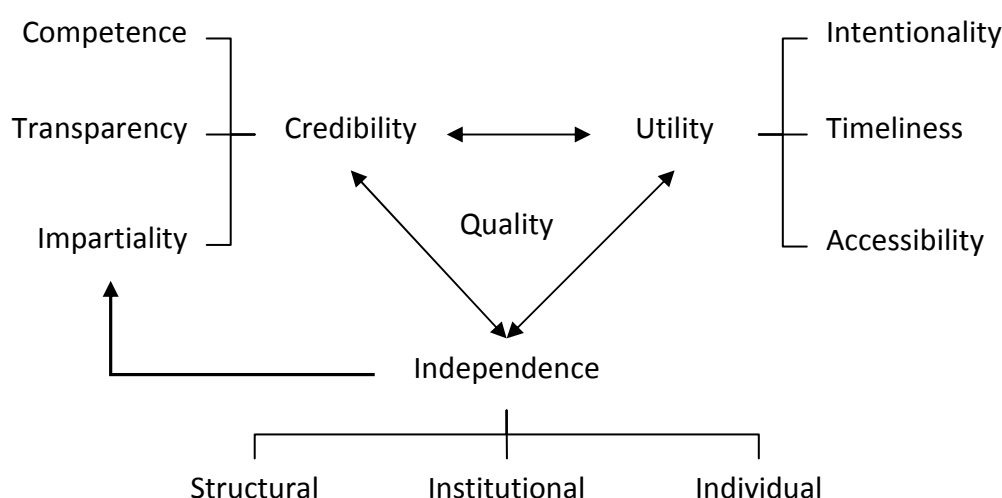
Independence forms the bedrock of good evaluation practice.⁶ It is fundamental to attain credibility and utility of evaluation. It should lead to impartiality reflected in, for instance, the choice of subjects of evaluation or the selected evaluation method. It is achieved through means at structural, institutional and individual levels.

Credibility of evaluation is enhanced with greater independence but needs to be accompanied by competence of the evaluators, transparency of the evaluation process, and impartiality of evaluators and process.

Utility of evaluation is not guaranteed with independence and credibility, but requires that commissioners and evaluators undertake the evaluation with the intention to use its result, undertake it at a time when the results can meaningfully inform decision-making processes, and that evaluations are accessible. This principle exists to ensure evaluations are undertaken to influence change so that governments and organizations can achieve their objectives and achieve them better.

Together these principles are markers of high quality evaluation and ensure “good practice” in evaluation. Independence, credibility and utility are three equal sides of the triangle at the centre of which the quality of evaluation rests. Their interrelationship, including inherent complementarities and tensions, is illustrated in Figure 1. Each of the central evaluation principles has other, supporting ones.

Figure 1: Evaluation Principles



Complementarities exist between these principles: an evaluation that is biased towards the perspective of one stakeholder (meaning: not independent or impartial) tends not to be credible with others. If an evaluation lacks credibility it is unlikely that stakeholders will pay attention to it, leave alone use it (meaning: its utility diminishes). Ignoring stakeholders, their interests or timelines, will reduce the utility of an evaluation but also the credibility of its process. By maintaining a good balance between independence, credibility and utility, these three principles will reinforce each other and enhance the quality of the evaluation.

But, tensions may also exist between the evaluation principles. For instance, independence can lead to isolation of the evaluation function which reduces its usefulness. Or some stakeholders might perceive evaluations to be useful solely when they report only successes

(and omit information about failures), or when they serve fund raising or publicity purposes. Or policy makers want to pursue a certain course of action that they do not like to see *derailed* by evidence that would advise otherwise. However, such evaluations have little credibility with other, well informed stakeholders.⁷ Equally important, they are of little, if any use for resolving problems. When problems remain unknown or covered up, solutions cannot be found to them and stakeholder will continue to be hampered in achieving their objectives. In other examples, stakeholders hold the belief that credibility can derive exclusively from close familiarity with the subject under evaluation, but ignore that such close link eliminates independence and reduces the chance that the evaluation will produce an impartial and credible evaluation that provides new insights.

Each evaluator must act in line with these evaluation principles.⁸ However, measures are needed to institute the independence, credibility and utility of evaluation to ensure adherence to these principles is less dependent on individuals alone. Evaluators may be subjective, lack an understanding that independence is granted to ensure impartiality and not for other reasons, and are at risk to be put under pressure from stakeholders who want to influence evaluation findings. Therefore, it is important to have an institutional framework that holds evaluators accountable and protects them from undue influence, and an enabling environment that supports the notion of evaluation, learning and accountability.

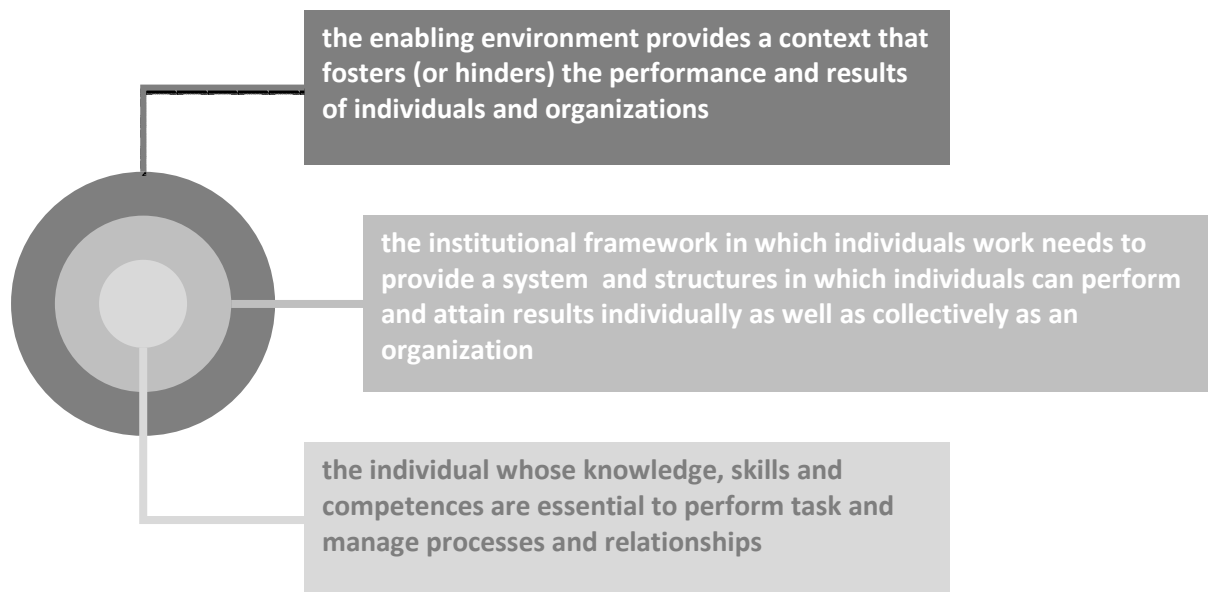
Capacity Development: Working at Three Levels

Capacity development is recognized by many as central to development and progress. The Accra Agenda for Action is the latest such commitment to strengthening capacities in support of ownership of development processes. Investments into developing capacities have been manifold, for many years and supported by many different actors, many of whom struggle with similar challenges. The concept of capacity development evolved from

- a focus on purely training of individuals,⁹ through
- institutional development, when it was recognized that individuals worked within the context of their organizations and therefore more than training was needed to be successful,¹⁰
- to capacity development to acknowledge that organizations do not work in isolation but require an enabling environment that consists of, among others, policies, networks, and an attitude of engagement.¹¹

Capacity therefore goes beyond an individual or an organization, but includes all three levels, as illustrated in Figure 2.

Figure 2: Three Levels of Capacity



Evaluation Capacities at Three Levels

Capacity development practitioners highlight the importance of working at three levels, which – when applied to the world of evaluation – illustrates the need to institutionalize the evaluation principles with measures that go beyond the individual, but span the institutional framework and the enabling environment for evaluation.

The ***enabling environment for evaluation*** is determined by a culture of learning and accountability, meaning the degree to which information is sought about past performance, the extent to which there is a drive to continuously improve, and to be responsible or accountable for actions taken, resources spent, results achieved. In such an environment, evaluation is understood to help decision-makers and implementers achieve common goals more efficiently and effectively. Such culture is embedded in tacit norms of behaviour, the understanding of what can and should – or should not – be done, in many cases behaviours being role-modelled by leaders.¹² They are, or should be codified in government legislation or an evaluation policy that express leadership’s or an organization’s commitment to learning, accountability and the evaluation principles. An enabling environment is also supported or created through governance structures that demand independent evaluation, be it through parliaments or governing bodies, and is further enhanced through professional associations and networks that set standards and strive towards greater professionalism in evaluation. Finally, the structural independence of an evaluation function is important to set an enabling environment: by placing the evaluation function in such a way that it does not report to the person or function responsible for the policies, strategies or operations that are evaluated, creates an environment that enables greater independence, credibility and utility. Ideally, the enabling environment is such that decision-makers proactively demand impartial evaluations to inform their debates and choices, which increases the usefulness of evaluations.¹³

The ***institutional framework for evaluation***

ensures that a system exists to implement and safeguard the independence, credibility and utility of evaluation. It reduces

- the risk that declared commitments to independence are revoked through personal actions: systems are tangible and, once established, are more difficult to reverse without agreement of all stakeholders; and
- the risk which is inherent in depending on the individuals and their personal behaviour, which can be supported through a system of checks and balances to ensure accountability and protection of the individual.

Such an institutional framework:

Includes a system of peer review or assurance that the evaluation function is set up to safeguard and implement the principles of independence, credibility and utility.

Establishes safeguards to protect individual evaluators – evaluators, evaluation managers, and heads of evaluation functions – when exercising their independence, including transparent and credible processes for the selection, appointment, renewal (if applicable) and termination of contract of the head of evaluation and assurances that evaluation staff do not suffer in promotion exercises or other repercussions.

Ensures a multidisciplinary team exists in the evaluation office and on any evaluation team that ensures credibility of the evaluation by understanding multiple dimensions of the issues and subjects and combining the necessary technical competence.

Secures the independence of funding of evaluations, at an adequate level, to ensure that necessary evaluations are carried out and that budget holders do not exercise influence or control over what is evaluated and how. Therefore, funding should be under the direct control of the head of the evaluation function, and should be adequate for an adequate work programme. The adequacy of funding for evaluation can be determined through a ratio of evaluation resources (human and financial) over the total programme size (some organizations aim to allocate 1-0.8 percent of their total budget for evaluation) or the coverage rate of evaluations (for instance, evaluating 25 percent of the overall programme, or to evaluate a minimum of 30 projects per year to have a valid sample size).

Structural independence requires that an evaluation function or office does not report to the same individual who is responsible for the subject of evaluation. Otherwise, the evaluation function might be under political or organizational pressure that interferes with the independent and impartial planning, conduct, and reporting of evaluation findings. The examples range from

- *National systems where evaluators report directly to Parliament to ensure the highest degree of independence and ensure this decision body benefits from impartial evaluation insights;*
- *Many of the development banks, starting with the World Bank, where evaluation functions report directly to the executive boards of these banks*
- *The UN system for which the Evaluation Norms suggest that a reporting line to the governing body and/or the head of the organization are adequate to ensure structural independence*
- *The practices of the Food and Agriculture Organization or the United Nations Development Programme, which introduced dual reporting lines: one functional reporting line to the governing body and an administrative line to the head of the organization.*

Structural independence does not guarantee impartiality. It requires the personal or behavioural independence of evaluators, and institutional measures that help safeguard independence and ensure it leads to impartiality.

Combines measures for impartial or purposive selection of evaluation subjects to ensure impartially on the one hand and increased utility on the other by making deliberate choices linked to decision-making processes.

- To avoid biases in the selection of subjects for evaluation, objective selection criteria are used to ensure the evaluation sample is representative of the whole. This principle would apply to the sample selection in a project evaluation when choices are made for site visits, the selection of projects or programmes for inclusion in an evaluation work plan of an organization or of initiatives of a national programme. As a result, the findings of the evaluation(s) will be as representative as possible, and therefore indicate strengths and weaknesses that are likely to apply to the project, programme or portfolio as a whole. There is no bias to select only the good performers or only the problem cases.
- To ensure utility, for instance of strategic, policy or thematic evaluations, these may be chosen purposively (rather than to establish a representative sample) to link the conduct of the evaluation to the information needs and decision-making processes of those using evaluation findings. In these cases, a consultation process for determine what are the most important/strategic topics for evaluation is important to ensure various stakeholder needs are considered.

Sets out a system to plan, undertake and report evaluation findings in an independent, credible and useful way. To increase objectivity in the planning and conduct of evaluation, systems are needed that increase the rigor, transparency and predictability of evaluation processes and products. Such systems can include more or less detailed process descriptions or guidelines for the design of evaluations, for preparatory work that needs to be conducted and reported on, and for reporting findings. The processes should have built in steps for communication, consultation and quality assurance. They should be communicated to stakeholders to enhance transparency and to secure their willingness to share information.

Institutes measures that increase the usefulness of evaluations, including the sharing of findings and lessons that can be applied to other subjects. Evaluations should be undertaken with the intention – of stakeholders and evaluators – to use their results. For projects or programmes this intention can derive from plans to use evaluation findings in the design in a future operation. Collectively the findings of these evaluations can be used to understand the performance of an organization with the intention to replicate well performing aspects and rectify systemic problems. The timeliness of planning and conducting evaluations and presenting their findings is equally important to ensure utility. Finally, accessibility means that evaluations are available to the public, can be retrieved (for instance through a user-friendly website), are written in ways that are understandable (clear language and limited jargon), and are distributed to a wide group of relevant stakeholders.

Even with structures and systems in place, the independence and impartiality of evaluation depends on the integrity and professionalism of **individuals** as evaluators, evaluation managers, and the managers of evaluation offices. The profession, and its ethics, requires limiting personal biases to the extent possible. In some circumstances, external evaluators are believed to exercise greater independence than those who work in an organization, because they are less exposed to institutional or peer pressures or have not absorbed the corporate culture in an unquestioning way.¹⁴ However, individual or intellectual independence **is** dependent on the individual, who demonstrates in his/her behaviour the

adherence to and practice of the evaluation principles: avoiding conflict of interest, acting with integrity and with an independence of mind, engaging in evaluations for which they are competent, acting impartially, and undertaking an evaluation with a clear understanding of the clients or decision-making process and how these need to be informed. Many evaluation associations and evaluation functions of national and international organizations adopted codes of conduct for evaluators and the UN Evaluation Group developed in addition ethical guidelines.¹⁵ Debates about professional standards and accreditation of evaluators and evaluation managers have been ongoing within professional forums for years.

To summarize Figure 3 provides an overview of the points discussed in the above paragraphs, illustrating the intersections between the three levels of capacities and the three evaluation principles.

Figure 3: Evaluation Principles and Three Levels of Capacity

	Independence	Credibility	Utility
Enabling Environment	<ul style="list-style-type: none"> ▪ Culture of Accountability and Learning ▪ Government Legislation or Evaluation Policy ▪ Community of Practice, Networks, Associations for Evaluation 		
	<ul style="list-style-type: none"> ▪ Governing or Oversight Body that seeks independent credible advice ▪ Structural Independence 	<ul style="list-style-type: none"> ▪ An attitude of providing access to information that allows credible evaluation 	<ul style="list-style-type: none"> ▪ An attitude that entails the intention to use evaluation findings and recommendations ▪ Willingness to change
Institutional Framework	<ul style="list-style-type: none"> ▪ Institutionalized process of “peer review” to assess the independence, credibility and utility of the evaluation function 		
	<ul style="list-style-type: none"> ▪ Independent Budget ▪ Impartial Selection of Evaluation Subjects ▪ Independent Planning and Conduct ▪ Non-interference in Reporting of Findings ▪ Measures to protect evaluators from repercussions 	<ul style="list-style-type: none"> ▪ Evaluation Quality Assurance System to ensure Credibility, Transparency and Impartiality ▪ Multi-disciplinary evaluation team that works well together 	<ul style="list-style-type: none"> ▪ Understanding of the use of evaluation ▪ Consultation processes ▪ Timeliness ▪ Accessibility of evaluations ▪ Active sharing of lessons from evaluation ▪ Utility demonstrated by the actual implementation of evaluation recommendations
	<ul style="list-style-type: none"> ▪ Ethics Guidelines ▪ Code of Conduct 		
Individual	<ul style="list-style-type: none"> ▪ Avoidance of Conflict of Interest ▪ Behavioural Independence ▪ Integrity 	<ul style="list-style-type: none"> ▪ Competence (technical knowledge and evaluation skills) ▪ Impartiality 	<ul style="list-style-type: none"> ▪ Client orientation ▪ Communication ▪ Coordination with different stakeholders and with ongoing processes

From Capacities to Capabilities: The Process Matters

Five Capabilities¹⁶

The above discussion centres around individuals and entities, such as governing bodies, management structures, evaluation units, or evaluators and how they collectively form an evaluation capacity of mutually reinforcing ingredients. To be effective, these evaluation capacities need to be capable of delivering evaluation services – producing independent, credible and useful evaluations – and much more. Otherwise, an evaluation policy might exist but is not acted upon, or a process is established but it is followed mechanically. A five year project on capacity development¹⁷ concluded that capacities needed to be capable

1. to commit and engage: *volition, empowerment, motivation, attitude, confidence;*
2. to carry out technical, service delivery and logistical tasks: *core functions directed at the implementation of mandated goals;*
3. to relate and attract resources and support: *managed relationships, resource mobilization, networking, legitimacy building, protecting space;*
4. to adapt and self-renew: *learning, strategizing, adaptation, repositioning, managing change;*
5. to balance coherence and diversity: *encourage innovation and stability, control fragmentation, manage complexity, balance capability mix.*

Evaluation Capabilities

This differentiation between capacities and capabilities is important. It moves the discourse from having, for instance, a written evaluation policy that however means little, because it is a document that is not backed by capabilities to commit and deliver against its promises. An evaluation unit may exist and may carry out evaluations, but might lack the capability to relate to stakeholders and attract resources, or to adapt and self-renew by learning to provide new evaluation types in response to changing needs.

In addition, these capabilities need to be interpreted differently depending on the stakeholders concerned. For instance, the commitment and engagement of stakeholders in the enabling environment will set the culture for learning and accountability, whereas the capability to commit and engage for an evaluation unit entails that it sets and abides by evaluation principles, and for the evaluator it requires an attitude of behavioural independence, among others.

The application of these capabilities to the three levels at which evaluation capacities need to be established is summarized in Table 1.

Table 1: Evaluation Capabilities

Capabilities	Enabling Environment for Evaluation	Institutional Framework for Evaluation	Evaluators, Evaluation Managers, Heads of Evaluations Units
to commit and engage	A culture of learning and accountability that empowers individuals and organizations to reflect on their practice, take stock of	The commitment to evaluation principles, and the motivation and attitude to follow them through in daily practice.	Commitment to evaluation principles, ethical guidelines and code of conduct. Behavioural attitude of independence

Capabilities	Enabling Environment for Evaluation	Institutional Framework for Evaluation	Evaluators, Evaluation Managers, Heads of Evaluations Units
	what works and what doesn't, and take necessary action.		and impartiality. Motivation to conduct evaluations in a credible and useful way.
to carry out technical, service delivery & logistical tasks	An evaluation policy that codifies evaluation principles and good practice.	Ability to conduct I/C/U evaluations to support organizational adaptation and renewal.	Competence to conduct evaluations, both from an evaluation and a technical point of view.
to relate and attract resources & support	To legitimize evaluation through the evaluation policy and actions, to recognize and ensure the need for adequate and independent funding.	The ability to demonstrate legitimacy based on impartiality and credibility, to protect evaluation principles and standards and the evaluators, and to generate adequate funding.	Capability to conduct evaluations transparently and credibly, and to communicate and relate with stakeholders.
to adapt and self-renew	Recognition of evaluation playing a role in adaptation and self-renewal.	The ability to make strategic choices about evaluation's engagement in organizational renew and to relate evaluation findings to stakeholders to ensure learning. Frequently adapt and update evaluation methods and approaches.	Learn new evaluation skills.
to balance coherence and diversity	Manage change, using evaluation evidence and recommendations when available. Avoid fragmentation of systems for learning and accountability.	The ability to ensure evaluation processes and approaches are systematic but flexible, balance capabilities on the team, employ a mix of approaches to develop an evaluation culture.	Find the right balance between systematically applying evaluation guidelines and seeking opportunities for innovation.

The enabling environment for evaluation needs to be committed to a culture of learning and accountability, adopts an evaluation policy – that is in line with the evaluation principles – and legitimizes evaluation, and uses evaluation findings and insights in its policy-making, performance improvements and organizational renewal. It accepts that the independence of evaluation needs to be safeguarded, including that of its funding. In a national context, this could entail embedding evaluation into legislation and government policy. In an organization, the corporate culture, evaluation policy, and resources for evaluation form the context in which an evaluation function operates.

The evaluation function should be committed to the evaluation principles, protect these and the evaluators from pressures, and have the motivation to implement them in their daily practice. It should have a system that safeguards and institutionalizes independence, credibility and utility. In its relationships with others it needs to demonstrate legitimacy based on impartiality and credibility (including technical competence) and it needs to contribute to the adaptation and renewal of the context within which it works – an

organization or a wider, say, public sector reform beyond just one organization. It has to have the capability to work systematically (to ensure transparency) but flexibly (to ensure credibility and utility), use a mix of tools to foster an evaluation culture, and keep its methods and approaches updated. The evaluation unit needs to have a mix of capabilities on its teams and has to have the capability to determine and negotiate adequate funding.

The individual evaluators – as staff of an evaluation unit or as consultants contracted to conduct a particular evaluation, as evaluators, evaluation managers or heads of evaluation units – has to be committed to the evaluation principles, the ethical guidelines and the code of conduct. S/he has to demonstrate independence and impartiality, and conduct evaluations in a credible and useful way. Evaluators are expected to be technically competent, as evaluators and/or in the field that is being evaluated, capable to deliver the evaluation service in question, and keep updated in his/her knowledge of evaluation methods and techniques. S/he should have the ability to follow guidelines, but seek possibilities for innovation.

The Process Matters

The process of capacity development is, in itself, an important part of the outcome, because capacities and capabilities are developed through a process of interaction and dialogue, mixed with specific initiatives to set up systems, install hardware, and train people. The process combines parts that are well planned, based on participatory diagnostics of capacities and their weaknesses, and others that are flexible and opportunistic. Just the number of stakeholders and the range of capacities and capabilities discussed above illustrate the complexities of capacity development and flag the need for planning and implementation tools that are able to address complexity, something that *blueprints* tend not to be good at.¹⁸ Instead, space needs to be created for analyzing capacity gaps in a participatory way, agree on common goals and translate them into a joint strategy while at the same time maintaining the flexibility and capability to take up opportunities as they arise, learn from experience and change tactics and work on various levels of capacity at the same time.

Capacity development practitioners arrived at the conclusion that ten process matters are important for greater success in capacity development. Here these process issues have been re-grouped around three themes: the drive from within, developing and agreeing on clear expectations, and “ordered chaos” or combining high degrees of flexibility with systematic approaches.

The *drive from within* has to do with the realization that capacity development needs to come from within and/or have strong internal champions, whether it is when developing a national system or a function in an organization. If an idea is imposed from the outside, chances are that capacity development is not likely to succeed. The *drive from within* relates to the capability “to commit and engage” and includes (1) ownership, which manifest itself in having local or internal champions, resource allocations, and engagement. It is, however, difficult to measure, is not homogenous across internal stakeholders, and is not constant over time; (2) leadership which is important, especially for setting the culture of learning and accountability, but it also must fit with the context and its culture; and (3) collective action, motivation and commitment, meaning that capacity development will not occur if it is one person writing documents, but requires engaging various stakeholders whose capacities and capabilities will be developed.

The process of *developing and agreeing on clear expectations in capacity development* is in itself part of the capacity development process. Developing a common understanding of capacity weaknesses through a participatory, structured diagnostic – using tools like the conceptual framework in figure 3 or the five capabilities in table 1 can serve as an analytical framework to move attention from resource gaps to recognize broader issues – and a shared vision of the needed capacities are important steps in channelling resources towards clear objectives. Very often the many actors in capacity development have tacit understandings of what capacity is and how its development should happen. Few of them have explicit strategies, which makes it more difficult to find a common ground and work in the same direction. A diagnosis of existing capacities is necessary to find entry points to start capacity development. This diagnosis benefits when it is developed in a participatory way that develops ownership at the same time.

The “*ordered chaos*” or *combining high degrees of flexibility with systematic approaches* entails a recognition that capacity development, while systematic, does not (always) follow an ordered process. Instead it includes planned as well as incremental and emergent approaches. For instance, when developing specific skills, a well structured process such as the ten steps¹⁹ would be appropriate. On the other hand, when trying to influence culture to become more supportive of learning and accountability, opportunities might arise at short notice, in various settings and without prior notice, and include an informal conversation with a key stakeholder. In these contexts, capacity development initiatives would be incremental and emergent in response to opportunities as they arise. Capacity development also requires recognizing the less tangible aspects of capacities and combining small and large initiatives, depending on context and opportunities. It requires finding the right balance between an operating space that allows capacities to evolve and accountability for capacity development results; a balance that is hard to attain. And finally, capacity development takes time, so it is necessary to stay the course (even through adverse times), but build *quick wins* into the process (to keep up motivation and build on success), and time to reflect and evaluate whether progress is being made.

Where to start? Illustrative examples: what does it mean for stakeholders?

There is no perfect answer when asking the question “where to start”. Depending on the situation, a diagnostic could be the best entry point to understand where the strengths, weaknesses and gaps are. Sometimes, however it might be necessary to first have a champion within who provides ownership and leadership of the process and generates interest in developing evaluation capacities. In other situations, it might be clear that the quality and credibility of evaluation has to be improved before anything else can happen. Or in other cases, the incentives for accountability and learning need to be improved before anything else can happen. And, in many cases it requires a mix of all of these things to develop capacities, but these choices are situation-specific.

An entry point for developing institutional capacities is a diagnostic, which can be done using the conceptual framework suggested here, and through peer reviews in the bilateral system, the United Nations Evaluation Group and Evaluation Cooperation Group of the multilateral development banks.²⁰ This type of expert review can support a diagnostic process and lend it further credibility, but the evaluation unit concerned needs to be involved in and own the diagnostic process. The diagnosis could lead to a revision of policy,

as has happened in the case of the United Nations Development Programme, and/or to the formulation of a strategy to develop evaluation capacities.

Very few organizations have articulated an evaluation strategy, even less so one that explains how evaluation capacities will be developed. Yet, good practice in capacity development argues for articulating such strategy to have a common vision and bring stakeholders together to work towards a common goal. Such strategy would employ *linear planning* approaches for some activities, for instance when implementing a training programme, *incremental approaches* when for instance gradually expanding the evaluation requirements in a national system as discussed in the Sri Lanka case study at the IDEAS conference in Johannesburg in 2009 to *emergent approaches* when a number of flexible, informal initiatives are needed to, for instance, to respond to opportunities when possible and necessary.

The following paragraphs highlight examples, many of which found in other chapters of this book, to draw out suggestions for stakeholders and how they might work towards developing evaluation capacities. It is structured around the three levels – the enabling environment, the institutional framework, and the individual – as these three levels associate well with identifiable stakeholders. At any of these levels, it is advisable to understand the existing capacities, the capabilities and reluctances to adopting certain standards to develop a corresponding and relevant capacity development strategy.

Placing Evaluation Capacities in the Context of Good Governance

The relationship between governance and evaluation is interdependent. Good governance creates an enabling environment for evaluation, while evaluation reinforces good governance.

In the book where this article will appear, the Sri Lanka case study provides a good example of how the commitment from the highest political level – the prime minister – created an enabling environment for evaluation. By demanding reports on results, stakeholders in ministries were enabled, even required to record and critically reflect on performance. This strong leadership was combined with a strong champion to translate political commitment into practical action.

Ideally, governments or chief executives of organizations seek independent, credible and useful advice through an evaluation function. Wiesner argues that demand for evaluation creates a strong enabling environment for evaluation. In this case, decision-makers demand feedback on the use of resources and results achieved with them and set, through their demand for such information, an environment that enables evaluative thinking and practice. However, Wiesner also suggests that political and institutional power structures can limit the environment in which evaluations are conducted, especially if vested interests resist evidence that might demonstrate the weaknesses of political choices. In these cases, it is important to counter-balance political power structures with other stakeholders and by establishing legislation or an evaluation policy that legitimizes and commits the country or organization to the evaluation principles.

An increasing number of United Nations organizations are adopting evaluation policies. In the case of the World Food Programme, the evaluation policy establishes the commitment to the evaluation principles and explains measures to safeguard them.

An environment that is unresponsive to or even fearful of evaluation might originate in a lack of understanding the value of evaluation to make more informed decisions. In these cases, it is for the evaluation function to deliver credible, high quality evaluations to demonstrate the value added of evaluation. In addition, it will be necessary to explain the role and usefulness of evaluation in the context of governance and performance of the organization as a whole. Another chapter in this publication provides an example where evaluation is embedded in the strategic management processes to increase the usefulness of evaluation, while decreasing negative perceptions. Or an incremental approach is illustrated in another chapter of the book to gaining confidence and creating an environment that eventually trusts the value of evaluation.

Tips for stakeholders to help reinforce an enabling environment for evaluation

Governing Bodies (Parliaments, Executive Boards, etc.)

- Be aware of the threats to independence, credibility and utility of evaluation and demand for measures to safeguard these principles
- Adopt and oversee the implementation of legislation and/or policies that institutionalize the independence, credibility and utility of evaluation
- Exercise oversight over the quality of evaluation
- Demand for evaluation and demonstrated use of evaluation findings and recommendations

Head of State, Minister, Policy Makers, Chief Executives

- Demonstrate leadership in setting a culture of learning and accountability
- Seek and use evidence from evaluations to validate the attainment of goals and objectives and to improve performance whenever possible
- Understand evaluation as part of good governance that aims to use public resources effectively and efficiently to achieve the goals that governments or organizations aim to achieve

Evaluation Function

- Demonstrate value added of independence, credible and useful evaluations
- Raise awareness of stakeholders about the role and importance of evaluation and of the evaluation principles
- Contribute to evaluative thinking through awareness building, dialogue, and training

Professional Evaluation Networks, Associations, etc.

- Set standards as benchmarks that can be used to convince other stakeholders about the importance of the evaluation principles and measures to safeguard them

Delivering Evaluation Services Independently, Credibly and Usefully

The importance of developing an institutional framework for evaluation – rather than using ad hoc arrangements – lies in the safeguarding of the evaluation principles, the provision of a framework that protects and holds evaluators accountable, and the predictability and transparency of this framework for all stakeholders. Such a system needs to be built in ways that it has the capability to adapt and self-renew, one of the five capabilities, to respond to

new challenges with, for instance, a focus on new issues or by introducing new types of evaluations that address information gaps.

Generally, it is the evaluation functions that need to lead efforts to develop the institutional framework for evaluation, setting up systems, processes and guidelines that are in line with the evaluation principles and fit with the work of their organization or the national context. Figure 3 above summarizes dimensions that an institutional framework for evaluation could or should entail and the rationale for including them.

These systems may include human resource management measures that ensure the selection of an independent and qualified head of the office of evaluation. The best systems will not – and should not – eclipse the human factor, but this makes it important to select and appoint heads of evaluation functions that are capable of safeguarding the evaluation principles and managing the evaluation function so that it delivers independent, credible and useful evaluations. The process should assure governing or oversight bodies and chief executives alike that an impartial candidate has been chosen and the selected candidate should have contractual arrangements that protect him/her from repercussions, should critical evaluations be unwelcome. In addition, it is important that a system is set up by which evaluators are held accountable for their actions through a code of conduct or ethics guidelines, such as those established by the United Nations Evaluation Group.

To set up a planning system that is representative for evaluations of operations, the Office of Evaluation at the World Food Programme analyzed the portfolio to identify factors that should be used to ensure a selected sample of operations is representative of the total. This resulted in a system that combines the size and number of operations to reflect the considerable variation in these factors: a relatively small number of operations represent a large proportion of the financial value of the portfolio, while the rest is spread over a large number of small operations. Treating each operation, regardless of size, equally would have overemphasized the smaller operations in the sample; selecting the sample by size would have under-represented the smaller operations in the overall sample. The weighted numbers are applied to determine how many operations in each geographical region should be selected for evaluation. By contrast, a different system was needed for the

Selecting the head of evaluation

1. The process should be agreed between key stakeholders
2. The process should be documented and publicized for reasons for transparency.
3. The composition of the panel needs to ensure credibility in the eyes of key stakeholders, requiring the representation of stakeholder groups and the profession.
4. The panel should have clear terms of reference that specify whether the panel “chooses” or “suggests”, and if it is the latter who has the final decision-making power.
5. The panel should be involved in the entire process from agreeing on the vacancy announcement, the selection criteria, agreeing on interview questions and criteria for assessing answers, interviews, etc.
6. Appointment should be made for a fixed term period, including a contract for the same duration.
7. The renewal of contract should follow a process that involves the panel to ensure the renewal is based on performance and not dependent on producing biased evaluations.
8. The termination process, while following standard human resource procedures, should again involve the panel to ensure termination of contract is based on performance issues and not a repercussion for critical evaluations.

strategic evaluations to ensure they address issues that are important to key decision-makers in the governing body and management, and revolves around a consultation process with these stakeholders. The utility of all evaluations is enhanced by linking them to decision-making processes.

An example of a system that ensures complete independence of finance is that of the International Fund for Agriculture Development. Its evaluation policy determines that the budget is prepared independently of that of the Fund's secretariat. The process is managed by the evaluation office and involves discussions and approval of the evaluation committee. The work programme and budget of the evaluation office appears as a separate, clearly identifiable section in the Fund's overall work programme and budget.²¹ This example can be contrasted by others, where funding of evaluations are dependent on the approval of project managers, who may or may not accept an independent evaluation.

A further system that is important to develop is one that governs evaluations processes and sets standards for evaluation products. They often take the form of guidelines or handbooks. Placing information about the process into the public domain increases transparency of evaluation processes. The production of guidance materials often needs to go hand in hand with briefings: between the evaluation manager and the evaluation team and between the evaluators and the managers of operations on the other hand to clarify expectations in evaluation quality and process. The process also needs to include formal feedback processes and quality standards and assurance mechanisms. If evaluations are conducted in a decentralized way, meaning: not managed by the central evaluation office, training will be needed to ensure standards are understood and implemented.

Evaluation functions need to have capacities to feedback findings and recommendations into decision-making processes as well as to share lessons from evaluations. These systems can include search engines on websites of repositories of lessons from evaluations, incorporating evaluation lessons into guidance materials that programme managers use, or more pro-active efforts to share evaluation lessons throughout the evaluation process. Many of these dissemination systems struggle with the challenge of providing lessons from evaluations to the communities where projects take place and where evaluators collected information.

Tips for stakeholders to help develop an institutional framework for evaluation

Governing Bodies (Parliaments, Executive Boards, etc.)

- Get briefed about the evaluation system to understand whether the institutional framework includes adequate checks and balances and to become a discerning reader of evaluation reports
- Request due process is followed for the selection, appointment, and termination of contract of the head of the evaluation function

Head of State, Minister, Policy Makers, Chief Executives

- Get briefed about the evaluation system to understand whether the institutional framework includes adequate checks and balances and to become a discerning reader of evaluation reports
- Introduce processes for the selection of the head of evaluation that ensures his/her independence

- Provide secure, separate and adequate funding for evaluation

Evaluation Function

- Develop and document systems for the selection, design, conduct and reporting on evaluations
- Provide briefings on these standards to increase transparency and confidence in the process and products, which enhances credibility
- Develop mechanisms to ensure lessons from evaluation are systematically shared and integrated into debates and decision-making processes.

Professional Evaluation Networks, Associations, etc.

- Set professional standards and good practice standards
- Develop and implement professionalization, accreditation, and credentialing systems

Impartiality, Knowledge and Skills of Individuals

The best enabling environment and greatest institutional arrangements will not replace the individuals involved in evaluation. Their integrity to plan, undertake and report evaluation findings in an impartial way is essential for the profession. They determine whether an evaluation is credible and useful. Their professional competences – as evaluators and in the technical area that they evaluate – and their ability to listen, understand and weigh the different views of stakeholders are central to the quality of evaluation.

The importance of training as a tool for one part of the capacity development process is well understood. The chapter on training approaches in this book illustrates the importance of a structured training programme that is based on specific training needs assessments and training goals. In addition, it is important to coach and mentor evaluation managers and evaluators especially when they are working for the first time in this field and/or work in an environment that is not enabling or supportive.

Tips for stakeholders to help develop skills and knowledge of evaluators

Evaluation Function

- Invest in training
- Build teams that are multi-disciplinary
- Encourage team work and cross-fertilization

Providers of Evaluation Training

- Continue to develop and offer training courses
- Increasingly move towards university courses to build a professional tradition

Professional Evaluation Networks, Associations, etc.

- Provide opportunities for professional exchanges and networking

Conclusion

This chapter demonstrates the synergies that exist between the evaluation profession and that of capacity development practitioners. A combination of good practice from both sides promises to result in greater effectiveness in developing evaluation capacities that will safeguard and use the evaluation principles of independence, credibility and utility and promises higher returns on investments in evaluation capacity development. It comes at an important time when evaluation, as a profession needs to professionalize, and when demand for evaluation and for developing such capacities is high.

The consideration of three levels of capacity (the enabling environment, the institutional framework and the individual) are important to institutionalize evaluation capacities in ways that it supports and protects individuals – evaluators, evaluation managers and heads of evaluation functions – while holding them accountable. These three levels, together with the capabilities that are needed, provide a conceptual framework that moves the debate from training individuals for improved skills or calls for resources to a more sophisticated set of ingredients that promise to be mutually reinforcing capacities.

Applying the conceptual framework to specific examples provides a number of pointers for stakeholders who want to work on evaluation capacity development.

Definitions

Accessibility means that stakeholders have unrestricted access to evaluations that can be retrieved easily from their storage place. It also requires that these evaluations are written in ways that are easily understandable and that findings are shared with stakeholders and interested parties.

Accountability is the obligation to account for (and report on) work carried out and results achieved.

Competence of evaluators means that they have – and can demonstrate – a good understanding of the subject under evaluation and of the evaluation principles, rigorous data collection and analysis, including ethical principles for evaluation.

Credibility is the extent to which evaluation findings and conclusions are believable and trustworthy. Credibility is determined by objective factors, such as the transparency of the evaluation process and accuracy of the evaluation report, and subjective factors such as the perceived or demonstrated impartiality and competence of the evaluators.

Learning means that lessons are drawn from experience and, accepted and internalized in practices, thereby building on success and avoiding past mistakes.

Impartiality is the absence of bias; it entails due process, methodological rigour, and the consideration and presentation of achievements and challenges, success and failures.

Independence means that the evaluation is free from influences – political or organizational pressures or personal preferences – that would bias its conduct, findings, conclusions or recommendations. It implies that evaluations are typically carried out or managed by entities and persons free of the control of those responsible for the design and implementation of the subject of evaluation.

Intentionality is a clear intent to use evaluation findings. In the context of limited resources, the planning and selection of evaluation work has to be carefully done.

Transparency means that consultation with the major stakeholders is an essential feature in all stages of the evaluation process. This improves the credibility and quality of the evaluation. It can facilitate consensus building and ownership of the findings, conclusions and recommendations.

Timeliness means that evaluations are chosen and undertaken in a manner that their findings and recommendations are available in time to inform decision-making processes.

Utility of evaluations means that they aim to and do affect decision-making. Therefore, evaluations must be perceived as relevant and useful and be presented in a clear and concise way. Evaluations are valuable to the

extent to which they serve the information and decision-making needs of intended users, including answering the questions posed of the evaluation by its commissioners.

Sources: *EC Evaluation Glossary (online)*, *DAC Glossary of Key Terms in Evaluation and Results-based Management 2002*, *DAC-UNEG Framework for Professional Peer Reviews 2007*, *UNEG Norms 2005*, *UNEG Ethical Guidelines for Evaluation 2007*, *World Food Programme Evaluation Policy 2008*

References

Arnold, Mary, Developing Evaluation Capacity in Extension of 4-H Field Faculty, AEA, 2007.

Asian Development Bank

Evaluation of Capacity Development Assistance to Lao PDR, 2004

Evaluation of Capacity Development for Managing Water Utilities, 2003

Evaluation of Country Assistance Program Evaluation: Mongolia, 2002

Evaluation of Technical Assistance for Public Expenditure Management to Bhutan, India, Kiribati and Lao PDR, 1999

Evaluation of Effectiveness of Training in Education Projects in Indonesia, 1997

AusAID, A staged approach to Assess, Plan and Monitor Capacity Building, 2006

Boyle, Richard and Lemaire, Donald (Editors), Building Effective Evaluation Capacity, Lessons from Practice, 1999

Capacity.org

Exploring the *soft* side of capacity development, 2005

Meso-level capacity development, 2004

Renewing approaches to institutional development, 2004

Capacity Development: the why's and how's, 2003

Capacity for Development: insights and innovation, 2002

Tools of the Trade: Capacity Assessment, 2001

Informing the capacity debate: operational experiences, 2000

Chelimsky, Eleanor, Clash of Cultures, *American Journal of Evaluation*, Vol. 29, 2008.

CIDA, Capacity Development: Why, What and How, 2000

DfID, Capacity Development and State Building: Issues, Evidence and Implications for DfID, 2005

Danida

A Results-Oriented Approach to Capacity Change, 2005

Capacity Development Evaluation Step 1: Contributions to an analytical framework, 2002

ECG

Review Framework for the Evaluation Function in Multilateral Development Banks, 2009

Template for Assessing the Independence of Evaluation Bodies, 2007

ECPDM

Capacity Development: between planned interventions and emergent processes, 2009

Capacity Change and Performance, Insights and Implications for Development, 2008

Study on Capacity, Change and Performance, Final Workshop Report, 2006

The Concept of Capacity, Peter Morgan, 2006

Inter-Agency Standing Committee, Capacity Building for Emergency Responses, 2007

ODI, Capacity Development for Policy Advocacy, 2006

OECD/DAC

The Challenge of Capacity Development, Working Towards Good Practice, 2006

DAC Principles for Effective Aid, 1992

UNEG

Ethical Guidelines for Evaluation, 2007

Code of Conduct, 2007

Norms and Standards, 2005

UNDP

Assessment of 20 Capacity Development Tools, 2005

Measuring Capacities: An Illustrative Catalogue to Benchmarks and Indicators, 2005

Capacity Development Action Briefs

Capacity Development Practice Note, 2007

Unicef, Capacity Building for Unicef Humanitarian Response in Review, 2002

UNIDO, Report on the Evaluation of UNDP-financed and UNIDO-executed projects of Technical Assistance in Metrology, 1990

World Bank

Capacity Building in Africa, an OED Evaluation of World Bank Support, 2005

Evaluation Capacity Development, 2006

Ten Steps to a Results-Based Monitoring and Evaluation System, A Handbook for Development Practitioners. Kusek & Rist, 2004

World Bank Institute

Capacity Development in the World Bank Group: A Review of Non-lending Approaches, 2007

A Revolution in Capacity Development? Africans Ask Tough Questions, 2006

World Food Programme

Evaluation of WFP's Capacity Development Policy and Operations, 2008

Notes

¹ Peter Morgan, in *The Concept of Capacity (2006)*, suggests that the capacity development is not well defined, no courses are given, and the practitioners do not form a coherent group. Nonetheless, the author of this paper argues that a professionalization of capacity development professionals is much needed.

² The Challenge of Capacity Development, Working towards good practice, OECD/DAC, 2006.

³ The Meeting was held by the evaluation network of the Development Assistance Committee (Organization of Economic Cooperation and Development). The evaluation principles were included in the "*DAC Principles for Effective Aid*" publication of 1992.

⁴ Many of the principles discussed then had also been discussed at and were later integrated into international and regional evaluation associations, such as the American and African Evaluation Associations (AEA and Afea, respectively), the Evaluation Cooperation Group of the multilateral development banks, or norms of the United Nations Evaluation Group. Sometimes the terms used are not identical, but the underlying principles are.

⁵ Framework for Professional Peer Reviews, DAC/UNEG Joint Task Force on Professional Peer Reviews of Evaluation Functions in Multilateral Organizations, January, 2007.

⁶ Chelimsky, among others, discusses in her article *Clash of Cultures* how the lack of independence affects the choice, design and methodologies used in an evaluation, all of which affect – adversely – the impartiality of the evaluation.

⁷ As Boyle put it "[e]ducated consumers can help create an evaluation "ethos" where evaluation is valued as an integral part of the government decision-making process.

⁸ Self evaluations, which are a valid evaluation tool, obviously are not as independent as evaluations undertaken or managed by an office or people not directly involved in the design, implementation and management of the operation.

⁹ Many evaluations still observe that capacity development places a heavy emphasis on training with limited understanding of the needs for other measures.

¹⁰ Report on the Evaluation of UNDP-financed and UNIDO-executed projects of Technical Assistance in Metrology, September 1990. The report defines "institutional capability" to include services to be performed, demand for these services, capabilities (skills in necessary quantity and quality) to meet demand, hardware (including facilities, premises and instruments), methodology, legislation, and management and coordination functions. These features, considered necessary for an institution to function, were integrated into project design training and the annual evaluation reports of UNIDO at the time.

¹¹ The need for an enabling environment was noted, among others, in the work of the European Centre for Development and Policy Management.

¹² Toulemonde speaks of building an evaluation culture as follows "[o]nce this culture is well established, evaluation is deeply rooted in the administrative values, is seen as an undisputed duty and becomes one of the fundamentals of the governing system. The culture provides the collective pressure that makes decision-makers overcome their reluctance, even when evaluation deeply contradicts their self-interest."

¹³ Mayne et al suggests "that to ensure utility of evaluations to various interests, structure and controls are not sufficient. Where they can have an influence, institutional users of evaluation must be proactive and assume responsibility for getting the kinds of evaluations they want others to produce."

¹⁴ Sonnichsen discusses, in particular in box 2.1 of his article, the advantages and disadvantages of internal and external evaluators and of centralized and decentralized evaluation units.

¹⁵ UNEG Ethical Guidelines for Evaluation, 2007.

¹⁶ The term "capability" in this context does not refer to Amartya Sen's *Capability Approach*, which is focused on the capabilities of the individual. Instead, it is used in the sense of "the ability to do or accomplish something".

¹⁷ Capacity Change and Performance, Insights and Implications for Development, ECDPM, 2008.

¹⁸ The Challenge of Capacity Development, Working Towards Good Practice, OECD/DAC, 2006.

¹⁹ See Kusek & Rist, Ten Steps to a Results-Based Monitoring and Evaluation System, A Handbook for Development Practitioners. Washington DC, World Bank, 2004.

²⁰ Frameworks for these peer reviews have been issued by the joint task force of the OECD/DAC and the UN Evaluation Group, and by the Evaluation Cooperation Group of the Multilateral Development Banks.

²¹ The work programme and budget can be found at (<http://www.ifad.org/evaluation/whatwedo/wp/index.htm>).