

International Initiative for Impact Evaluation



WORKING PAPER 8

Institutionalisation Of Government Evaluation: Balancing Trade-Offs

Marie M Gaarder
Bertha Briceño
July 2010

About 3ie

The International Initiative for Impact Evaluation (3ie) works to improve the lives of people in the developing world by supporting the production and use of evidence on what works, when, why and for how much. 3ie is a new initiative that responds to demands for better evidence, and will enhance development effectiveness by promoting better informed policies. 3ie finances high-quality impact evaluations and campaigns to inform better program and policy design in developing countries.

3ie Working Paper series covers both conceptual issues related to impact evaluation and findings from specific studies or synthetic reviews.

This Working Paper was written by Marie Gaarder, 3ie and Bertha Briceño, World Bank.

© 3ie, 2010

Contacts

International Initiative for Impact Evaluation
c/o Global Development Network
Post Box No. 7510
Vasant Kunj P.O.
New Delhi – 110070, India
Tel: +91-11-2613-9494/6885
www.3ieimpact.org

INSTITUTIONALISATION OF GOVERNMENT EVALUATION: BALANCING TRADE-OFFS

Marie M Gaarder

International Initiative for Impact Evaluation

New Delhi

Email: mgaarder@3ieimpact.org

Bertha Briceño

World Bank

Washington DC, USA

Abstract

Carefully designed and implemented evaluations can improve people's welfare, and enhance development effectiveness. The paper investigates institutions in Mexico, Chile, and Colombia, and shows that for the successful inception of an institutionalised system for evaluation, three common factors stand out: the existence of a democratic system with a vocal opposition, the existence of influential M&E champions to lead the process, and a clear powerful stakeholder. Mexico's CONEVAL is the most independent of the three bodies, mainly due to the fact that it is reporting to an executive board of independent academics; Chile's Dipres is the best placed in terms of enforcement, with its location within the Ministry of Finance and control of an independent budget; and Colombia's Sinergia helps promote a culture of utilization of evaluations as a project management tool. However, actual usage of M&E information and the resulting effect upon development effectiveness are the benchmarks of success. The paper concludes that an explicit and thoughtful process of assessing the needs, the focus, and the emphasis of the system should serve officials and champions to identify adequate arrangements for the particular country context and understand how to better respond to the forces pushing for the creation of new M&E units and bodies.

I. Introduction

Policy-makers are experimenting with billions of people's lives on a daily basis without informed consent, and without rigorous evidence that what they do works, has no substantive adverse effects, and could not be achieved more efficiently through other means. Non-evaluated policies that are being implemented are by far the most common experiments in the world. Nevertheless, parliaments, finance ministries, funding agencies, and the general public as citizens and tax-payers are starting to realize this and are demanding to know how well development interventions achieve their objectives, not only whether the money was spent or the schools built. In this context, carefully designed and implemented evaluations have the potential to save lives and improve people's welfare.

However, to date evaluations have tended to be selected based on the availability of data, the interest of researchers and donors, the amenability to certain evaluation methods, and the availability of funds rather than on their potential contribution to broader development strategies. This paper discusses the rationale for institutionalising government evaluation efforts, and the main considerations and trade-offs that have to be made, drawing on existing experiences from Latin America.

II. Monitoring versus evaluation

In this paper we focus on government evaluation and the monitoring thereof. Monitoring and evaluation, also commonly known as M&E, are terms that tend to get mentioned in one breath, yet although the activities are related, the main functions they fulfil, the time-lines, the actors involved, and the sources of funding can be quite different, and it is the exception rather than the norm that both the monitoring and the evaluation of a specific activity are done under the same institutional arrangement.

While monitoring is used to continuously gauge whether the project or intervention is being implemented according to plan, evaluations assess progress towards and the achievement of outcomes (and possible unintended outcomes), and *impact* evaluation whether these can be attributed to the intervention. Monitoring is a continuous process, while evaluation should be done at a point in time when the project activities can be expected to have a measurable impact. Monitoring is usually done by implementation staff, while evaluation can be either internal or external, and similarly the users of the former type of information are mainly program managers whereas the latter is also used to inform the wider public, including parliaments, press, policy-makers, and the international community. Funds for monitoring are more likely to be an intrinsic part of a program budget than for evaluation. Information which is collected from monitoring is useful for continuous program improvements, while information resulting from evaluations is available at a later stage and therefore more often used to improve the design of a new program phase, to make decisions regarding the survival or expansion of the program, or to inform policies in other settings (public good).¹ For all of these reasons, it is not surprising to find that most organizations do establish monitoring systems as an integral part of their activities, whereas evaluation tends to be more of an afterthought, often externally imposed.²

¹ Both monitoring and evaluation systems are most useful if they are incorporated into a program or intervention from its inception, however in the case of evaluation a number of techniques allow for evaluations to be realized later in the program life.

² To complicate matters, however, the concept of evaluation encompasses a number of different methodologies, including consistency and results evaluation (a logframe type of evaluation), process evaluation, benefit incidence and targeting evaluation, beneficiary satisfaction evaluation, a range of qualitative evaluations,

In addition, because of the vested interests of program staff in the survival of a program, the lack of incentives to implement major changes (and even disincentive as it usually implies additional work), and the lack of distance enabling to see 'the forest rather than the trees', it is generally accepted that some form of external, more objective or independent entity needs to be in charge of the evaluation. As we will see in the following section, there are a number of reasons why that entity could be usefully one single entity in charge of the evaluation efforts of an entire public sector, or indeed public evaluation efforts more generally. Evaluation of this institution's activities in turn, should then be the subject of an independent evaluation. What then of the monitoring efforts? Indeed, as each agency should be continuously monitoring the implementation process and progress of its activities, so should an agency in charge of evaluations monitor the progress of its evaluation agenda.

So when we talk about institutionalising M&E, we need to be clear about what it is we are monitoring (the projects or the evaluations), what it is we are evaluating (the projects or the evaluation program), as well as what is meant by institutionalising.

III. Why institutionalise?

The term *institutionalisation* is used in social theory to denote '...the process of making something (for example a concept, a social role, particular values and norms, or modes of behaviour) become embedded within an organisation, social system, or society as an established custom or norm within that system'. However, the term may also be used '...in a political sense to apply to the creation or organisation of governmental institutions or particular bodies responsible for overseeing or implementing policy, for example in welfare or development' (Wikipedia; 07/06/2010).

Applying these definitions to the area of M&E, we would then distinguish between institutionalisation at the level of the implementing organisation, and institutionalisation at a more aggregate level, be it at a sub-sector (e.g. Ministry) or sector level or at a national public policy level. While it is generally accepted that project monitoring is a *sine qua non* for implementing organisations to continuously keep a finger on the pulse of the project and ensure efficient implementation and necessary course corrections (i.e. institutionalisation in the first sense of the term), the level at which responsibility for evaluation should lie is a more debated theme and is intrinsically linked to two distinct issues: (i) evaluation should be a *tool for policy-making* to ensure improvements in the allocation and effectiveness of scarce resources, and (ii) evaluation should seek to be *independent and relevant* to ensure credibility and usefulness.

We therefore understand institutionalisation as a process of channelling isolated and spontaneous program evaluation efforts into more formal and systematic approaches, on the presumption that the latter provide a better framework for fully realizing the potential of the evaluation practice. This is because, if we accept that program design adjustment, policy realignment and feedback into planning and budgetary processes are the *raison d'être* of program evaluation, it is sensible to believe that strategic orientation, rules and organisational immersion will make evaluations more policy-influential. At the core of the practice of evaluation is comparison and benchmarking, and analysis of trade-offs. As we argue below, policy-making is thus likely to be enhanced when a systematic approach enables to compare results across different interventions and intervention designs.

impact evaluations, and a host of others. Each of these draws on different data sources, and in particular draws on program monitoring data to a different extent. While an impact evaluation could in theory be carried out with minimal interaction with the program and program staff,² the process evaluation naturally has to be done in close collaboration with the same.

Influence in policy is more likely when independent bodies are in place to understand and channel the needs of evaluation's clients, and are able to define a strategic orientation according to them; thus enhancing relevance.

A. Policy-making tool

Using evaluation to achieve optimal allocation of resources requires knowledge not only of the impact of interventions, but also of alternative uses of the funds. Knowledge of the outcomes of certain programs will inform policy-makers whether the program or intervention is indeed contributing to the achievement of the results it was set out to achieve, which combined with the budgetary information informs about the cost of this achievement. Knowledge of the impacts of policy alternatives on the same outcomes and the related costs will be able to address relative cost-effectiveness of the interventions.

Although perfectly sensible in theory, in practice a number of factors complicate the picture. First of all, it is rare that programs have identical objectives, outcome measures and target populations, making a cost-effectiveness comparison difficult. Second, the opportunity costs are related to an infinite amount of other possible programs, in all sectors, not just the sector in question, making optimal allocation based on evaluation results virtually impossible. Moving away from the ideal of optimal allocation, however, information on the benefits (or monetized value of the impacts) and costs can be used to answer whether the allocation is acceptable in the sense of bringing returns at a level policy makers deem acceptable (e.g. IRR above 12%).

Nevertheless, there are two additional very good reasons for making a comprehensive review of ongoing programs and interventions an integral element of policy-making. First, it alerts central authorities to non-functioning programs (i.e. not delivering on their intended outcomes) and to areas for improvements, and, second, by linking the evaluation process to the budget process offers central authorities an instrument to enforce evaluation activities and the implementation of recommendations. To fulfil these objectives, the central entity tasked with overseeing the evaluation process needs to have a structure that has the ability to prioritize, the ability to set standards for methodologies and practices, and the authority to influence policy with the outcomes.

B. Independence and relevance

While oversight requires central level involvement, independence and how to achieve it is a more debated theme. Within the auditing and evaluation communities one generally distinguishes between *independence of mind* (the state of mind that permits the provision of an opinion without being affected by influences that compromise professional judgment) and *independence in appearance* (the avoidance of facts and circumstances that are so significant that a reasonable and informed third party would reasonably conclude that integrity had been compromised), noting that both are closely linked (IFAC, 2010; 21). The Glossary of Key Terms in Evaluation and Results Based Management issued by the Development Assistance Committee of the OECD specifies that an evaluation is independent when it is "carried out by entities and persons free of the control of those responsible for the design and implementation of the development intervention". It also indicates that independent evaluation presumes "freedom from political influence and organizational pressure", "full access to information" and "full autonomy in carrying out investigations and reporting findings" (OECD, 2002; 24).

Independence is only one dimension of evaluation excellence, and without relevant skills, sound methods, adequate resources and transparency, quality is not guaranteed. Furthermore, it is important to note that optimum independence is unlikely to be full independence, since some relationship with the implementing agency is usually needed to

ensure relevance, access to information for conducting the evaluation, and influence of recommendations. 'The ability to engage with diverse stakeholders and secure their trust while maintaining the integrity of the evaluation process is the acid test of evaluation professionalism' (EES, 2008: 2). So, while external evaluations tend to be equated with independence, their relevance is often diminished by their lack of appreciation of the operating context and access to operational information. Furthermore, the reality of their independence is determined mainly by who is funding them, and may be compromised if it is by the very managers in charge of the activities under evaluation. No undue influence needs to be exerted by the managers for the situation to be compromised, as consultants may be self-censoring to maintain their clients. Internal evaluations, on the other hand, while clearly more likely to be influenced by internal politics, can in principle be partly shielded from undue management influence if they are funded and controlled by an autonomous governance entity (IADCI, 2008; 5).³

Even when an evaluation may achieve credibility by having been performed with independence and quality, this does not per se ensure objectivity in the reporting of the ensuing results. If the organization in charge of reporting the findings either to the public or to central authorities is the one in charge of the activities under evaluation, then reporting objectivity and independence is compromised, at least the appearance thereof. If on the other hand there is a law or regulation in place prescribing the dissemination of all evaluation documents and results or if the organization reporting is free from organizational pressure from the one overseeing the activities being evaluated, then independence and credibility are more likely to be ensured.

There is however a third level at which independence may be compromised, even when organizational independence of the reporting entity is assured, and that is by political influence. If the entity overseeing the public evaluation efforts is vulnerable to political changes (e.g. if its existence is politically determined, its budget allocation, or its staff) then the independence in mind and appearance principles may be compromised, as it may be or feel under pressure to report successes only.

However, a location outside of the executive while more likely to achieve independence may come at a cost: the downside of the completely external arrangement is that as the system becomes more separated from internal budget or planning authorities, its power to enforce or exert direct influence over the objects of oversight may be less direct. Transparency and accountability utilisation might be stronger, at the expense of utilisation as an internal management and control tool from the government's centre (budget central authority, planning, presidency or internal control office).

To summarise, while it is clear that a central governmental institution or particular body in charge of overseeing the public evaluation efforts is necessary if evaluation should be a tool for overall policy-making, in line with the second definition of *institutionalisation*, it is also clear that the credibility of the information reported by this agency will depend in part on its independence while the enforcement capability of improvements to the programs and to the allocation of the national budget will depend on proximity to the government centre.

The following section will discuss the three leading models and experiences of national evaluation bodies in Latin-America, from Mexico, Colombia and Chile, highlighting first how they came to be created, and subsequently how each has dealt with the trade-off between independence and influence.

³ Four interrelated dimensions of evaluation independence have been recognized by the Evaluation Cooperation Group (ECG), including: (i) organizational independence; (ii) behavioural independence; (iii) protection from external influence and (iv) avoidance of conflicts of interest.

IV. Balancing trade-offs in three Latin-American cases⁴

A. *Inception*

A conjunction of factors in the early 2000s cleared the way for the institutionalisation of evaluation in **Mexico**. Among these factors were an increasing demand and technical assistance for evaluation from multilateral agencies, mainly the Inter-American Development Bank and the World Bank, as well as the appearance on the program scene of an innovative program for poverty alleviation, known as PROGRESA (later Oportunidades), which incorporated rigorous evaluation as an integral part of the program from the outset (see box 1). This program in particular, and the evaluation agenda more generally, was promoted by certain evaluation champions, including the influential Mexican economist Santiago Levy, then serving as Deputy Minister at the Ministry of Finance and Public Credit. Among the enabling factors, the possibly single most important one in the creation of the central evaluation entity, the National Council for the Evaluation of Social Development Policies (CONEVAL⁵), was the strong political pressures from the opposition, culminating in the enactment of the 2004 Social Development General Law, by which the evaluation process was institutionalised.⁶

The institutionalisation of evaluation in **Colombia** was related to a historical process leading up to the 1991 constitution, by which the country signed a new social agreement emphasizing the participatory character of the democracy and the role of social control. The constitution, and Law 152 of 1994, explicitly assigned to the National Planning Department, NPD,⁷ the mandate for promoting evaluation and performance-based management in the public sector. A second factor that contributed to the institutionalisation was the fact that after the experience with the evaluation of the Mexican conditional cash transfer program, PROGRESA, the multilaterals were pushing strongly for the evaluation of social programs. Accordingly, a social safety net was also launched in Colombia in 2000, the so-called *Red de Apoyo Social* (RAS), which included 3 social programs identified by multilaterals as promising projects to be evaluated. Funds from the loans were thus earmarked to carry out independent evaluations. Another important factor that allowed resurgence of the evaluation system after a period of stagnation during the late 1990s was the endorsement that President Uribe's first administration gave to the management for results culture.

The evolution of the management control system has been a long-standing effort of the **Chilean** government under the leadership of successive budget directors. The origins of the system date back to the early 1990s, a period characterized by the consolidation of public reforms. The program of evaluation was launched in 1997, responding to a demand from Congress for further quality information and influence over budgetary decision-making. Indeed, the recently created International Advisory Panel for Evaluation and Management Control System have recognised that "the increasing emphasis on evaluation within the Chilean context has been in part in response to demands from Congress for more and better evaluations and for the increasing use of such evaluations to guide public resource allocations". As from 2000, the administration of President Lagos promoted a more integrated vision of state modernization, and created the management control division within the Ministry of Finance to implement the evaluation and management control system (World Bank, 2005; 30). In 2003, a formal legal mandate requiring evaluation of public programs was introduced (Dipres, 2008 a; b; and c).

⁴ This section draws on the 3ie report 'Institutionalising Evaluation: A Review of International Experience' (2009) commissioned by DFID.

⁵ <http://www.coneval.gob.mx>

⁶ Diario Oficial, México, 2004a; 2004b; 2005..

⁷ An administrative department with ministerial status,

While the historical particulars vary, the three stories around the inception of the institutionalised systems for evaluation have many common elements. The existence of a democratic system with a vibrant and vocal opposition appears to have been an important enabling factor, as has the existence of influential M&E champions to lead the process. A clear powerful stakeholder, such as Congress, the Ministry of Finance, or the Presidency facilitates triggering the process and an external incentive and push from multilaterals also was a common trait. Finally, the power of examples of influential evaluations, as was the case with Oportunidades, were important triggers (Box 1). Once constituted, however, how can these centralized institutions be maintained and made effective?

Box 1. An Influential Evaluation: Oportunidades

Mexico's Conditional Cash Transfer (CCT) program Oportunidades is a social protection program aimed at alleviating poverty in the short-term, while promoting human capital accumulation and thereby breaking the inter-generational poverty-cycle. CCT programs provide cash to poor households upon household compliance with a set of health and education-related conditions. Expected immediate results include increased food consumption, school attendance and preventive health care utilization among the poor. Longer-term expected impacts are increases in the accumulation of human capital and associated returns in the labour market.

The program started to operate in rural areas in 1997 under the name of Progresa. By 2001 it had been extended to semi-urban areas, and by 2002 it reached urban areas. Five million families currently benefit from this program; approximately 25% of the population and all the poor.

From the outset, an evaluation component was included to quantify the program's impact through rigorous methodologies (focused on attribution rather than contribution), using both qualitative and quantitative approaches. The work was assigned to internationally and nationally renowned academics and research institutions.

Perhaps the **largest impact** of the evaluation thus far, with very positive and credible results emerging, is its important role in ensuring that the program was not eliminated with the change of government, contrary to what had become the norm for previous changes in administration. The name of Progresa was however changed to Oportunidades to mark the change.

Another important impact, to which the Oportunidades evaluation experience has contributed, has been the adoption of a Mexican Law which now requires all social programs to have yearly external evaluations of their programs.

An external "impact" of the program has been that a number of other countries in the region have adopted similar programs to Oportunidades, including Colombia, Nicaragua, Honduras, El Salvador, Panama, Costa Rica, Paraguay and Jamaica.

Finally, a number of modifications to the design of the program have been made as a result of the evaluations, including (i) an extension of the education grants it provides, beyond junior high to the high school level, as the evaluation revealed larger program impact on schooling attendance of children of secondary school-age; (ii) improvements of the methodology used in the health talks, from a passive lecture-style to an interactive and more hands-on learning approach; (iii) adjustment of the health talk content to address the urban challenges related to chronic diseases, risky behaviour and unhealthy life-styles; and (iv) adjustment of the food supplement composition to include a type of iron that would more easily be absorbed.

B. Independence

As argued in previous sections, an oversight body should enjoy a high degree of independence to be able to freely make assessments and fully disclose them without improper influence. Presumably, the higher the degree of independence, the higher the credibility of ensuing findings and the better the reception from clients outside the government, such as Congress, the media, and civil society. Evidence to support this presumption which illustrates how varying degrees of independence have played out in practice, for example in the ability to publicise negative findings, is made difficult by the very fact that the results that are made publicly available would already have been through a censorship process if such exists. A comprehensive study interviewing researchers involved in evaluations in different systems, to gauge the degree of censorship they experience and at which points during the evaluation design, implementation and reporting phase, would be extremely useful and has to our knowledge not been carried out to date.

Appearance of independence is first and foremost associated with the organizational location, with institutions positioned outside of government assumed to enjoy a higher degree of independence. Nevertheless, there are other factors that can influence the independence of an organization. In the following discussion of the Mexican, Chilean and Colombian institutionalisation efforts, we will distinguish between *organizational location, source of funding, reporting structure* and *dissemination laws* when analyzing the degree of independence of the evaluation oversight bodies. In addition, we will distinguish between oversight bodies that are also in charge of commissioning and supervising the external evaluations, and those that leave this mainly to the agency under which the activities to be evaluated take place.

In 2000, impelled by a Congress mandate, the **Mexican** government began to measure poverty and evaluate its social programs for the first time. Measurements obtained indicated that poverty was decreasing, and that social programs were successful, but the opposition strongly mistrusted these results arguing that they were own statements lacking objectivity. As a result, CONEVAL was established⁸ with a twofold mission: to measure poverty (national, state and municipal level) and to ensure and oversee the evaluation of all social development policies and programs at the federal level to improve results and support accountability practice under methodological rigor. Although the mandate of CONEVAL is formally constrained to the social sector, it acts as the standard setter and articulator of evaluation activities across government agencies. Different units within each ministry or sector agency carry out evaluation activities at various degrees, under the guidance and coordination of CONEVAL.⁹

Despite the original demand by the opposition to locate CONEVAL outside of government, it was in fact placed under the Ministry of Social Development, but with technical and managerial autonomy, including a head appointed directly by the executive. The potential compromise to their independence due to the possibility for exertion of political pressure from the said Ministry was however in part counteracted by two factors; first of all, CONEVAL's operating costs (though not the evaluations) is financed through a direct budget line in the National Budget; and second, it is governed by an executive board of six independent academics.¹⁰ This board is appointed by the National Commission for Social Development, a commission made up of representatives from the federal states, municipal representatives, delegates from Congress and the executive, tasked at consolidating and integrating social development strategies and databases.¹¹ Identification of candidates for the six positions is managed through a public bidding process.¹²

A general law introduced in 2002 of access to public information, was further operationalised in CONEVAL's General Guidelines prescribing the dissemination of all evaluation documents and results through the internet websites of the relevant department or entity within 10

⁸ Diario Oficial, México, 2005.

⁹ The broader picture of government M&E activities comprises other institutions that perform monitoring and auditing activities at the central level. Those practices are more aligned with performance based management practices. They are basically monitoring and budget execution follow-up activities led by the SHCP, and auditing activities carried out by the SFP. There are ongoing initiatives to create units of evaluation under each of these institutions. Three areas can therefore be identified where an institutionalisation gap remains in Mexico: (i) the alignment of central evaluation efforts between these new evaluation units and CONEVAL; (ii) the lack of evaluation at the sub-national government levels; and (iii) the relative absence of institutionalised evaluations (impact evaluation and other, such as process evaluation) in the non-social sectors.

¹⁰ The Board also includes the Minister of Social Development, and the Executive Director of CONEVAL.

¹¹ It comprises 32 officials from social development entities at the federal level; the heads of the Ministries of Social Development, Education, Health, Labour, Agriculture, the Environment and Natural Resources; a representative from each of the national municipal associations; and the presidents of the Social Development commissions in the Senate and Chamber.

¹² Criteria for members include being or having been members of the national system of researchers and having broad expertise in the subject of evaluation or poverty measurement.

business days of their reception.¹³ The mandated dissemination helps ensure the transparency and objectivity of the evaluation reporting process. Most of the evaluations carried out under this system are however commissioned and supervised by the agencies in charge of the activities under evaluation rather than by CONEVAL directly, something which makes the evaluation reports vulnerable to biases before they get published.

Thus, the reporting structure of CONEVAL and the public dissemination law ensures the institution a degree of independence and immunity from the current political regime, and the direct budget-line and autonomous status within the Ministry of Social Development ensures it a degree of independence from the said Ministry. However, the evaluation reports it receives from the social sector federal agencies may suffer from biases before reaching CONEVAL.

In 2000, **Chile**'s administration under President Lagos consolidated the evaluation and management control instruments within the budget department of the Ministry of Finance, Dipres¹⁴. The overall goal of the unit is to contribute to the efficiency of allocation and utilization of public spending, contributing to better performance, transparency and accountability. The evaluation of programs line of work includes governmental program evaluations (1997), impact evaluations (2001), and the evaluation of new programs (2009). The latter emphasise inclusion of evaluation at the design stage and including control groups when possible. Dipres has the technical support of an International Advisory Panel of renowned professors in the impact evaluation field which gives recommendations regarding the technical design of evaluations of new programs, the necessary data collection, and supporting the process and the results analyses.

The definition of the evaluation agenda is closely linked to the budgetary annual cycle, and is supported by Congress through the signature of a protocol in November every year for selected programs to be evaluated. The source of funding for evaluations in the protocol is Dipres' own budget line. Agencies may fund additional evaluations and establish other monitoring instruments through their sector budgets. The evaluation plan is shaped and approved by an Inter-Sector Committee, which is chaired by a representative of the budget directorate, and includes representatives from the Ministries of Finance, Planning and of Secretary of Presidency, but the main influence is exerted by Dipres (Mackay, 2007; 27).

The head of the management control division reports directly to the Budget Director under the Minister of Finance. The Budget Directorate is accountable to the Congress which has a say in the approval of the protocol of selected programs to be evaluated (it can request the inclusion or removal of certain programs or institutions within the annual evaluation plan). Seemingly, the Congress has not been very active in modifying the evaluation agenda (Rojas et al., 2005; 8). The evaluations of programs and institutions are reported to Budget, Congress and the public, and are available at Dipres' website. Also, in 2008 Chile introduced a Law of Transparency and Access to Public Information.

Thus, Dipres is clearly dependent on the Ministry of Finance both in terms of its organisational location and lines of reporting, and Congress has only a marginal role in counterbalancing this dependence. The main factors that may add to the credibility of the reported findings are therefore the commitment to public dissemination of reports, the existence since 2008 of the International Advisory Panel that advises on the quality of the

¹³ In addition, they mandate for internet disclosure of contact information of the external evaluator and the program responsible, the type of evaluation, databases, data collection instruments, a methodological note with description of the methodologies and models used along with the sampling design, and sample characteristics; an executive summary with main findings and the recommendations of the external evaluator; and finally, the total cost of the external evaluation, including the source of funding.

¹⁴ <http://www.dipres.cl/572/channel.html>

impact evaluation designs and processes for the evaluation of new programs, and the fact that Dipres itself oversees the external evaluations, rather than the agencies being evaluated.

In 1994, **Colombia** established SINERGIA¹⁵, the national system for evaluation of public policies and management for results. It is conceptualized as a national system so that it conveys a complete set of actors that are involved with monitoring and evaluation activities, and their roles. Such actors include providers of M&E services (academia, research centres, private firms and consultants), governmental agencies, plans, policies, and programs (as objects of M&E, recipients and users) and other producers and recipients of M&E information (statistical institutes, civil society organizations, congress, media). President Alvaro Uribe, elected in 2002, injected new life into the system by making SINERGIA a corner-stone in his results-based management approach to government.

SINERGIA's mandate and conceptual basis are broad and involve M&E activities across all sectors and government levels. In practice, the Directorate for Evaluation of Public Policies (DEPP) acts as the technical secretariat of SINERGIA. It is a unit established within the National Planning Department (NPD), a long-standing administrative department with ministerial status that acts as technical arm of the Presidency, coordinating and guiding policy-making along with sector ministries, and in charge of central government's investment budget. In practice, DEPP's main scope of action is related to its regular interaction with agencies and ministries at the central level regarding monitoring of the system of goals and ongoing evaluations of programs, capacity building activities and dissemination of M&E information.

As compared to CONEVAL, DEPP does not enjoy technical and managerial autonomy. DEPP is headed by a technical director, responding to NPD's deputy director and a general director, who have the status of Minister and Vice-minister, respectively.¹⁶ Furthermore, consultancy staff and dissemination activities are financed mainly through NPD's investment budget, thereby also creating a budgetary dependence on NPD.¹⁷ However, in an attempt to provide the system with a "whole of government" reach beyond the sole influence from NPD, an Inter-Sectoral Evaluation Committee (IEC) was established, chaired by NPD's deputy director and including representatives from the Ministry of Finance, NPD directorates, and principal sector ministries. The IEC was given the responsibility for overseeing the government evaluation agenda, in addition to coordinating evaluation processes, approving methodologies, and considering the results that may contribute to improving the formulation of policies. This committee has however to date functioned on an ad-hoc basis, with limited ownership and 'buy-in' to the evaluation agenda from its members, and there is no provision for an extra-governmental governance body, such as in the case of CONEVAL. More indirectly, DEPP/SINERGIA is answerable to the Presidency, as are all public agencies under the management for results framework.

The position of dependence of DEPP within the Ministry of Planning could have been partly remedied by the introduction of an external governing body, as the academic board of CONEVAL, and of clear public disclosure laws (as in Chile), both of which are currently lacking and would imply a broader legal reform. As for the question of who is supervising the external evaluations; it is sometimes DEPP (usually when the activities are being financed by loans from the multilaterals) and sometimes collaboration between

¹⁵ <http://www.dnp.gov.co/PortalWeb/Programas/SINERGIA/tabid/81/Default.aspx>

¹⁶ In practice, DEPP's head also reports in an ad-hoc manner to the Advisory Minister to the Presidency, as one of the main users of the M&E information provided.

¹⁷ Resources for evaluations come primarily from the programs; some evaluations have had support from multilaterals which earmark resources for evaluation within the loan budgets.

DEPP and the agency who is overseeing the activity being evaluated (mainly in the case of federal agencies who self-select into the collaboration).

So, while Mexico's CONEVAL scores better on independence than the comparable bodies in Chile and Colombia, what has this entailed in practice in terms of the quality of the reporting? There is some anecdotal evidence to suggest that the degree of independence is related to the echelon at which any type of censorship occurs. Given that CONEVAL enjoys a relatively high degree of reporting independence, but is usually not in charge of commissioning and supervising the evaluation studies makes the latter an area susceptible to undue influence by self-interested parties. There is anecdotal evidence to suggest that CONEVAL has had difficulties getting an insight into the evaluation processes in some of the federal agencies reporting to it. In the case of DEPP in Colombia, on the other hand, where there is little reporting independence, where the agency is closely involved in the actual commissioning and quality-assurance of the flagship studies, and where the survival of the agency in large part depends on the continued demand from governmental and multilateral agencies for their services, there is evidence indicating that the visibility and dissemination efforts by DEPP related to the evaluation reports are censored and determined in part by what was politically useful, rather than the other way around, where political decisions are based on the findings. Often, this meant that more positive reports were given more visibility, or that decisions are independently made of the findings.¹⁸ Finally, in the case of Chile, where the evaluation agenda, commissioning, supervision and reporting of evaluations are in the firm grip of the Dipres within the Ministry of Finance, the susceptibility to bias may lie in the fact that the Ministry determines which programs get to be evaluated. Furthermore, until recently the quality of the reporting suffered from methodological limitations and lack of quality filters, but the existence of the International Advisory Panel is bound to help rectify this situation.

C. Policy Influence

The gains from being 'outside' of government can come at a cost. As the evaluation system becomes separated from budget and planning authorities, it may have less power to enforce or directly influence the adoption of recommendations by the implementing organisations, and by the planning and budgeting authorities. In this sense, presumably, location within budget authorities provides the strongest powers to the system to enforce adoption of recommendations derived from the assessments, thus ensuring utilization. In some cases, laws that make evaluation compulsory for inclusion by budget or planning authorities, or formal requirements to respond to recommendations and implementing them, can act as substitutes for having direct institutional access to these authorities. Furthermore, central evaluation bodies with access to own financial resources to carry out the evaluations also enjoy more enforcement capability.

As we saw in the previous section, none of the three cases is located outside of government. Indeed Dipres in Chile is located within the budget authority, as close to enforcement power as is possible, with a dedicated budget line to finance the approved evaluation plan. In the case of Mexico, CONEVAL's enforcement capability over the social sectors does not lag much behind, given that the social sector agencies are required by law to have an annual evaluation program agreed-upon with CONEVAL, the Ministry of Finance (*Secretaría de Hacienda y Crédito Público*, SHCP), and the public comptroller's office (*Secretaría de la Función Pública – SFP*) as a prerequisite for inclusion in the national budget. DEPP in Colombia has neither an institutional location nor the backing of a law to give teeth to its evaluation oversight mandate. In the cases of both CONEVAL and DEPP, the resources for

¹⁸ In the case of an urban work-fare program, *Empleo en Acción*, the decision of closing the program was prior to the evaluation results (indeed, the evaluation was nick-named 'the autopsy'), and in the case of a youth training program, *Jóvenes en Acción*, it was completely transformed before the availability of results, in spite of substantial positive effects found afterwards.

major evaluations come primarily from the program budgets, rather than their own, reducing this avenue for control.

The strongest enforcement capacity is hence clearly in Chile, and this is also reflected in the fact that the Chilean system's M&E information is highly utilized in budget analysis and decision making, in imposing program adjustments and to report to the Congress and civil society. One of the strengths of the Chilean system is that it maintains very specific information regarding program changes and monitoring of recommendations derived from evaluations. Given that the standardized terms of reference for the evaluations ensure that very specific recommendations are prepared, these serve as a basis for establishing Institutional Commitments (*compromisos institucionales*) which afterwards are closely monitored by Dipres. However, managerial usage or ownership from the head of the programs has been limited, given the centrally-driven nature of the system and the perceived absence of incentives for the agencies to engage in their own evaluations. Some shortcomings with respect to the quality of the findings have also been evidenced in the past, most probably due to the limited budget allocated to evaluations and the ex-post nature of the same (Mackay, 2007; 29)

The risk of low enforcement capabilities can be addressed in diverse ways to ensure that the evaluation efforts feed into policy-making. Support from Congress, fluid communication, and promotion of alliances with government central authorities are common strategies to mitigate weak enforcement of recommendations. CONEVAL's alliance with the Ministry of Finance and DEPP's alliance with the Office of the Presidency of the Republic are examples of these *de facto* channels for influencing policy.¹⁹

An alternative strategy to promote the adoption of recommendations is generating a tradition of utilization as a managerial tool rather than a control tool -- persuasion as opposed to imposition. If the implementing agencies are involved in identifying the issues to be addressed by the evaluations, and consulted in the design, implementation and analysis phases, then a sense of ownership of the evaluation efforts may ensue which will also increase the likelihood of utilization and voluntary adoption of recommendations by the program managers. To achieve this type of voluntary uptake, and program demand for evaluation, the central evaluation body is required to invest highly in demonstrating the benefits of evaluation as a managerial tool, in capacity building activities and in providing guidelines and tools.

SINERGIA is the prime example of this latter approach. Given its demand-driven orientation and limited enforcement powers, DEPP's focus has been on the utilization of evaluation information by program managers. DEPP is generally recognized as the agency with the technical expertise to support the various agencies in their IE endeavours. It provides advice on methodologies, support in the construction of TOR, as well as manages some evaluations. It also provides technical advice and financial support for some of the sophisticated impact evaluations conducted by sector ministries and agencies. It has experience in bidding processes, negotiation expertise with the evaluation firms and knowledge of the evaluation market and costs. Over the years, these services are powerful incentives to make the ministries and agencies turn to DEPP when interested in carrying out impact evaluations, building up legitimacy.

The ownership of the evaluation process by program implementing agencies is arguably due in part to a self-selection bias, whereby agencies more open to evaluations will approach DEPP for collaboration,²⁰ but also to the approach the institution takes to

¹⁹ Monitoring information is used extensively by the President and his office as a control tool.

²⁰ This approach will tend to favour 'stronger' programs and institutions, leaving perhaps those most in need of evaluation the possibility to opt out.

dissemination and the adoption of recommendations. For each evaluation, the institution carries out an intensive and step-wise dissemination process, starting with technical staff, continuing with the managers and heads of units of the program under evaluation, and finally with the heads of the agencies, the respective minister, the budget director (MOF), the President's Advisor with ministerial status, and the General Director of DNP. Externally, DEPP has organized seminars and events for academia, government, and policy-makers, where the external firms are invited to present the evaluation, and each presentation is followed by a discussion with a panel of experts. This step-wise approach, with the incorporation of feedback at every level, minimizes the sense of unfair public exposure by program staff and managers. Documentation exists on the changes in the programs adopted as a result of each evaluation undertaken, and a new practice of ensuing action plans is being implemented. The downside is the limited use from budget authorities and Congress, as well as the reluctance from civil society and media to acknowledge impartiality.

A second line of activity within SINERGIA is the system for performance indicators which tracks progress against the president's goals, SIGOB. DEPP coordinates the reports of sector ministries and agencies, and sub-national governments, which provide the monitoring information needed for the SIGOB. This line of activity could have given the institution some leverage over the evaluation agenda, however, according to MacKay (2007) the agenda has so far been decided in a bottom-up manner rather than in a planned, top-down manner. In particular, the agenda is currently highly influenced by the international donors who include evaluation as part of their loans to the government, together with individual sector ministries more open to evaluation. If in the future SIGOB's performance information could be used to flag poorly performing government programs for which an evaluation could be warranted, then the system would become more relevant as a budget and planning tool for central government.

In the case of CONEVAL, the set-up of the evaluation system and guidelines is designed to address both managerial usage and budget and planning usage. In 2007, jointly with the Ministry of Finance (SHCP), and the Public Comptroller's Office (SFP), CONEVAL issued the General Guidelines for the evaluation of federal programs²¹. First and foremost, the guidelines stipulate that CONEVAL jointly with the Ministry of Finance (SHCP) and the Public Comptroller's Office (SFP) define an Annual Evaluation Program for the federal institutions of the social sectors. The purpose of the Guidelines is to help regulate the evaluation of federal programs and establish principles and requirements for the different components of the M&E system. They are mandatory for federal public administration dependencies and entities that are responsible of federal programs. In 2008, general procedures to track improvement aspects derived from the different types of evaluations was established, reflecting main findings, responsibilities, recommendations, and measures taken to improve the programs based on the recommendations.²² A technological platform for maintaining and updating this tracking system is being developed, hosted by CONEVAL. Hence, the agencies' commitments and progress are accessible and open to scrutiny by the public. Examples of recommendations identified as a result of this exercise include improving the targeting mechanisms of the federal programs, improving the effective coordination among institutions and programs, improving information systems of social federal programs, some particular recommendations for the education and health sectors, and some recommendations on measuring results and coverage (CONEVAL, 2008).

²¹ Diario Oficial, México, 2007.

²² Aspects to improve are classified into three types according to their nature: specific (those that are the responsibility of the program officers), institutional (those requiring attention from various units within the agency), inter-institutional (requiring attention of external agencies) or inter-governmental (requiring attention of different government levels). The sector agencies themselves classify the aspects as of high, medium or low priority, according to their perceived contribution to the achievement of the program's goal.

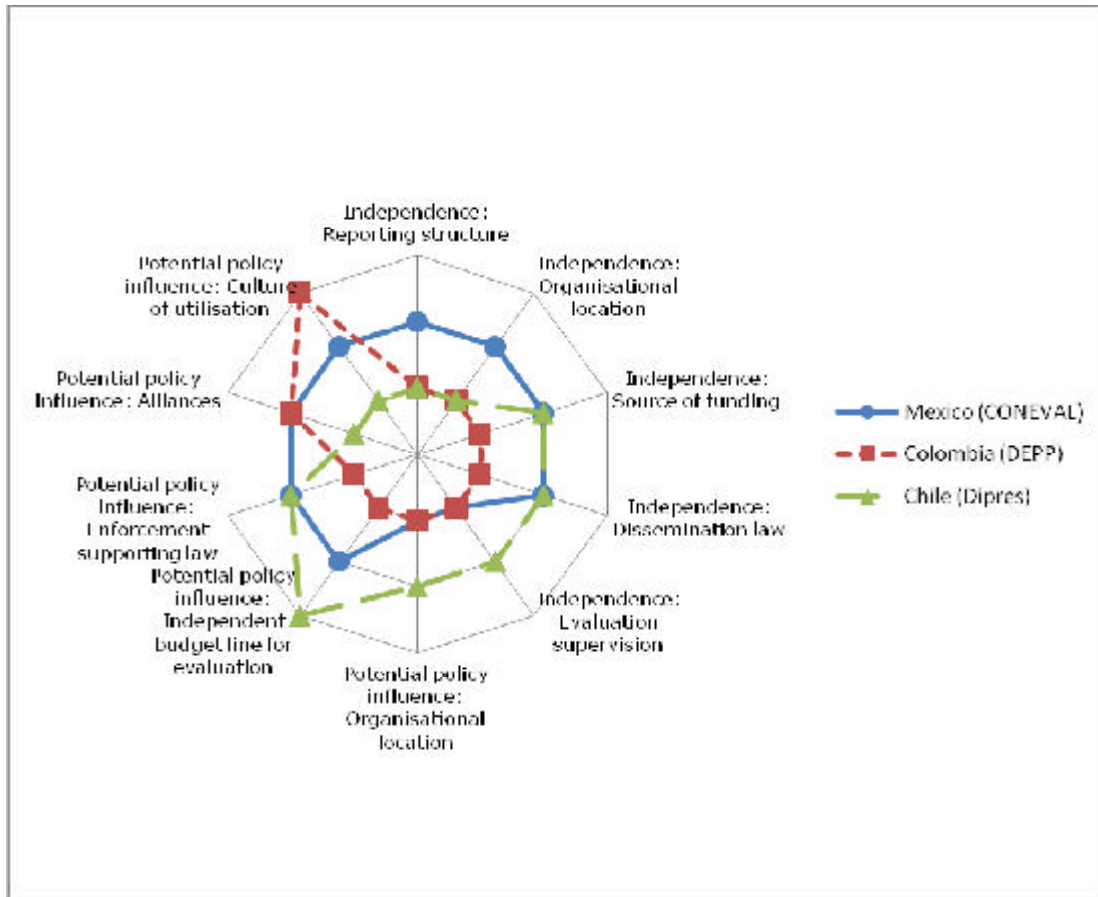
In order to achieve buy-in by program managers, CONEVAL also arranges training seminars and provides inputs on the suggested methodologies and Terms of Reference. In addition, it is recent practice that the officials who manage the evaluated programs have a say in which recommendations they deem actionable, and their performance is measured against the implementation of these agreed-upon actions. The main risk of this approach is that the implemented changes will be those that are marginal rather than larger changes, such as shutting down ineffective components of a program.

Figure 1 gives a visual summary of how the three cases fare in terms of their level of independence versus their potential for policy influence, with a higher score reflected by a larger distance from the centre, and the three country cases distinguished by colour. It is important to note that the scores have no numerical interpretation, nor are the scores comparable across the ten aspects included in the diagram. Rather, the diamond-shaped diagram shows the relative ranking of the three country-systems on each aspect, using the following criteria; (i) the lowest scoring system on each particular aspect has always been awarded a 1; (ii) if the two other countries score similarly on the particular aspect they are both awarded a 2; and (iii) if the three score differently on the aspect in question they are awarded a 1, 2, and 3 (with the system/country awarded 3 being the one that scores the best on that particular aspect). The following are the aspects that we have argued may contribute to independence, together with an indication of how to attain a high score in each case: (i) reporting structure – this area gives a higher score to systems that report to a body that is over and beyond the current political interests; (ii) organisational location – systems with some degree of managerial and technical autonomy score better on this aspect; (iii) source of funding – systems that have a direct budget line in the national budget to finance its operating expenses achieve a higher ranking; (iv) dissemination law – systems/countries with laws in effect that prescribe the publication of evaluation reports and follow-up commitments perform better on this aspect; and (v) evaluation supervision – systems in which the body in charge of supervising the evaluation is different from that in charge of the activities being evaluated score better here. As for potential policy influence, we distinguish between the following aspects: (i) organisational location – an evaluation body within the Ministry of Finance scores higher; (ii) independent budget line for evaluation – the existence of this gives the central evaluation body more direct control over what gets evaluated and when; (iii) enforcement supporting law – this refers to the existence of some law that makes it difficult or impossible for the federal programs to refuse evaluation; (iv) alliances – this refers to the existence of influential stakeholders that support the evaluation efforts (beyond the formal relations); and (v) culture of utilisation – this refers to a situation where program managers are persuaded rather than forced to do evaluations.

Overall, CONEVAL is the most independent of the three bodies, mainly due to its technical and managerial autonomy within the Ministry of Social Development, and the fact that it is reporting to an executive board of six independent academics appointed by the National Commission for Social Development. While the institution scores relatively higher on independence, it is usually the federal entities in charge of the activities under evaluation that supervise the studies, which constitutes a threat to the independence under which these are being performed. Chile's Dipres is by far the best placed in terms of enforcement, both due to its location within the Ministry of Finance and because it has control of an independent budget line to finance the evaluation plan that Congress approves. The main threat to sustained quality policy influence being the lack of ownership of the evaluation process by program implementing agencies. Finally, Colombia's system distinguishes itself by employing persuasion and dissemination strategies that help promoting a culture of utilization of evaluations as a project management tool. Threats to this system are two-fold; first, the credibility of reported findings is questioned due to the lack of independence and public dissemination laws, and second, the ability to enforce recommendations is lacking, thus making the system rely on voluntary adoption by those

agencies who are voluntarily submitting themselves to evaluation, thus introducing a potential double bias.

Figure 1: System trade-offs in Mexico, Colombia and Chile



V. Measures of success

While in the previous sections we characterised the systems based on how well they perform on aspects that theory and literature predict are important for well-performing central evaluation oversight bodies, the actual usage of M&E information is the benchmark of success, and determines the sustainability of the systems. Idiosyncratic developments and cultural features shape the focus of the M&E system utilization, resulting in distinct combinations from single to multiple clients and usages. We have identified clients from the Executive, such as Planning and Budget Ministries, which seek to improve the efficiency and effectiveness of resource allocation. Other clients include the implementing agencies which are generally more interested in revising implementation processes, fine-tuning the design, changing and improving managerial practices, and responding to its constituencies with concrete information. External clients include multilaterals and donors, Congress and civil society, with a focus on transparency and accountability purposes, as well as on broader lessons learnt.

As the saying goes, ‘the proof of the pudding is in the eating’. Directly, this will imply that programs that have acted upon the recommendations resulting from evaluation efforts have improved their performance on the desired as well as the undesired outcomes, measured through second generation evaluations. At the macro-level it will imply a continuous updating and revision of the priority outcomes, to ensure that the outcomes that are being improved upon remain sector and country-priorities. Indirectly, however, the findings from

evaluations of particular programs can have learning effects for other programs, even in different sectors or countries, and the culture of evaluation itself may have positive spill-over effects, implying that most direct measures of the effect of institutionalising evaluation upon development effectiveness may be biased downward. These types of measures of the impact of institutionalising evaluation are still lacking, and indeed establishing attribution will remain the biggest challenge. In the absence of these types of measures of success, what has typically prevailed are output and outcome measures that result from the evaluation bodies' monitoring systems.

Defining measures of success in terms of utilisation is not an easy task, and is an endeavour that the systems only recently are beginning to undertake more carefully. The World Bank (WB) has contributed with actively promoting some assessments of the systems' performance and diagnoses (Mackay; 2007, Rojas et al., 2005). CONEVAL recently commissioned an assessment of its General Guidelines for Federal Programs Evaluation from a World Bank team, another team carried out a comprehensive analysis of the Chilean public expenditure evaluation program in 2005 (Rios, 2007), and the Independent Evaluation Group (IEG) published a diagnosis of SINERGIA in 2007 (IEG, 2007). The *Centro Latinoamericano de Administracion para el Desarrollo* (CLAD) has continuously studied the systems since the late 1990s, and in 2006, engaged jointly with the WB in an ambitious initiative to strengthen the region's M&E systems. They used a standard methodology to analyse 12 countries, resulting in a series of individual country studies and a 2008 comparative report (CLAD-WB, 2008). So far, this can be considered the major and more significant effort to assess the evolution of the systems at the regional level. The CLAD-WB assessments involved case studies with structured interviews with the main stakeholders, potential and actual users, and staff responsible, whereas the Chilean World Bank evaluation included a revision of samples of evaluation reports, assessed comparatively against certain standard criteria. Sinergia's diagnosis was mainly a case study with in-depth interviews and documentation revision.

Two dimensions have been particularly explored in search for indicators of success of evaluation systems. First, what can be referred to as coverage, a measure of the extent of the evaluation activities in relation to a reference value or universe. Usually, the indicator would be either the proportion of the budget evaluated, i.e. the value of the programs that have been evaluated to the total budget amount, or the number of programs evaluated in relation to the number or programs in a programmatic classification of the budget. The second dimension refers to the utilisation of the evaluation results, and typically relates to tracking the commitments and action plans derived from the evaluations, as well as the follow-up of recommendations. This can be for instance simpler measures as the number of changes derived from evaluations and number of recommendations adopted, or more demanding ones, as the proportion of the recommendations implemented over the total number of recommendations formulated. Table 1 provides an overview of available indicators for coverage and utilisation.

Table 1 : Tracking performance of government based M&E systems

Coverage	Proportion of budget/programs evaluated	Budget of evaluated or monitored programs over total budget amount
		Number of programs evaluated or monitored over multi-year agenda
		Number of programs evaluated over number of programs in programmatic classification of budget
Utilization	Follow-up on recommendations, commitments and actions plans derived from M&E information	Number of changes derived from evaluations
		Number of alerts generated from monitoring
		Number & list of recommendations adopted
		Number of recommendations prioritized and adopted
		Number of recommendations implemented over total number of recommendations formulated
	Transparency/accountability	Number of incidences associating transparency or accountability with information from M&E systems
	Improving quality and efficiency of public expenditure	Number of the programs that have acted upon the recommendations resulting from evaluation efforts that have improved their performance in second phase evaluations
Changes in budget/resource allocations resulting from utilization of M&E findings by Congress		

CONEVAL reports figures related to both these dimensions, although the picture is incomplete. There are between 100 and 130 federal programs under the mandate of CONEVAL (reported figures differ by year), of which all are required to carry out logframe-type evaluations for which it provides Terms of Reference and guidelines. In addition, CONEVAL oversees directly about 15 evaluations per year, equivalent of 11% of the programs under its mandate, of which approximately 20% are impact evaluations. What is not clear is how many additional evaluations are taking place under the auspices of the individual implementing agency. For the 2008 budget exercise, 101 programs were included in the tracking system, with 930 aspects to improve. Out of these, 73% were included by 3 entities, and 70% were of the specific type (those that are the responsibility of the program officers) (CONEVAL, 2008).

The 2008 Public Finances Report by Dipres also presents measures of utilization. Between 2000 and 2008 approximately 174 programs were evaluated when taking into account the two traditional instruments of program evaluation, namely, the governmental program evaluations and the impact evaluations. Out of the total of programs, 27% were required to undergo a substantive program redesign, 37% required modifications in the design and internal management processes, 23% required minor adjustments, 6% recommended an institutional relocation, and 7% have been programs eliminated or completely replaced or absorbed. Regarding commitments, between 1999 and 2007 more than 3500 have been established, around 500 annually in the early years and lowering since 2006. Out of these, 82% were fulfilled, 11% were partially fulfilled, and 6% have not been fulfilled. The ministry of education is the entity with more programs evaluated (28) (Dipres, 2008).

An underexplored area to date is the assessment of the quality of the recommendations and action plans that emerge from the evaluation systems. Evaluators' main role is to identify areas within a program in need of improvement, but they are not necessarily the best placed to make specific recommendations. Nor is it clear that the implementers of the program have the required distance to the program or the incentive to identify the needed changes, or to pick which recommendations to follow up on, as they do in Mexico. An independent assessment by a panel of sector specialists of the

recommendations and action plans that ensue from the evaluations against the evaluation results would be one possibility to gauge the quality. Another way would be the second generation evaluation of the 'improved programs', as mentioned previously.

Measures in other dimensions, like transparency and perception of accountability by citizens, for instance, surveys exploring the connection or direct relationship between these and performance of M&E systems, or particular evaluation practices, have not been used to our knowledge.²³ In addition, when the system has also an orientation towards influencing budget allocations, further utilization measures could include the change in allocations as a result of evaluation utilization by budget and Congress or Parliament, or more indirectly, correlation measures with resources allocation changes.²⁴

To date, assessments of the success of systematised evaluation efforts have been limited to measures of evaluation coverage, clients' satisfaction surveys, some evidence on adoption of recommendations and commitments, and some anecdotal evidence. A more systematic collection, monitoring, and evaluation of the recommendations and commitments will be required to draw further lessons for the existing systems and for other countries starting out.

VI. Conclusions

We started out by proclaiming that carefully designed and implemented evaluations have the potential to save lives and improve people's welfare, and more generally be a powerful tool for development effectiveness. This paper reviews the experiences of institutionalising government evaluation efforts through a discussion of three leading models in Latin America – Mexico, Colombia and Chile, in an effort to provide a framework of characterization that enables to derive lessons for countries starting down that road.

We used as a framework for comparative analysis a core wish-list of features that, in theory, a best-practice M&E system should deploy. Overall, we want a system that is independent in order to achieve external credibility and social legitimacy, but not so independent that it loses its internal relevance. The placement of the system and the powers to publicly disclose the information produced, without a bias towards positive results, are key determinants of independence, credibility and legitimacy. It is important to enjoy a unique and broad legal mandate to ensure enforcement of recommendations, and avoid competing initiatives that undermine consolidation and legitimacy. Legal support from Access to Public Information or Transparency Laws is also an important asset to back full public disclosure, especially in systems located within the executive. We observed best practices such as the transparency laws and mandates of public disclosure in Chile and Mexico. In terms of independence, a best practice example is provided by external governing bodies like the academic board of Mexico's Coneval.

In addition, we want a system that is able to influence policy-making and the adoption of recommendations, either by promoting ownership or by using enforcement powers. This should not be a spontaneous but a purposeful process, defining clear channels built into mandates and preferably through legislative powers providing the evaluation body with a say in resource allocation. Chile's Dipres followed a strong strategy in terms of enforcement, both due to its location within the Ministry of Finance and because it has control of an independent budget line to finance the evaluation plan. Not having a

²³ Should they exist, though, confounding effects will need to be dealt with to actually give a sensible attribution to the effect of evaluation practices.

²⁴ For an interesting example on this potential measure, examining the correlation between evaluation results and budget growth of evaluated programs in Korea, see Kim and Park (2007) and Park (2008).

location close to the budget authorities or complete budget autonomy, Colombia's Sinergia and Coneval rely more on a combination of managerial buy-in and capacity building strategies, coupled with important alliances to foster influence. Sinergia has distinguished itself by its dissemination strategies, while Coneval provides an excellent example as a standard-setter.

Finally, we want a system that is sustainable over time and transcends governments because it is perceived as responsive to the needs of clients and useful to its main stakeholders. For this, the performance of the systems should begin to be tracked. Chile's Dipres provides a good example of this. Also, there needs to be a clear focus on usage and clarity on a client or set of clients that are to be served, and what their interests are. It can be Congress or Parliament, the broader society, central government or program management. Finally, fundamental to the production of, demand for and use of evidence/evaluations is the building of local technical capacity among relevant Ministry officials, program implementers, and local researchers, as well as the strengthening of data collection and processing systems in order to ensure high quality of data.

In terms of inception of an institutionalised system for evaluation, three common factors stand out from the cases discussed in this paper. Firstly, the existence of a democratic system with a vibrant and vocal opposition appears to have been an important enabling factor, as has the existence of influential M&E champions to lead the process. Furthermore, a clear powerful stakeholder, such as Congress, the Ministry of Finance, or the Presidency facilitates triggering the process. In addition, technical assistance and the existence or training of technical capacity in the country has been important enabling factors both for the inception and sustainability of the systems.

Thus it is clear that the *wish-list* of features can be sought and achieved through different evolution paths and that along such paths each system adopts particular choices and defines its own trademark. As the inception and evolution of the systems show, the underlying trade-offs in the focus and clients' orientation depend on the political and cultural contexts. Specific circumstances have shaped –and will continue to shape– the inception, evolution and focus of each system, and accordingly, its capacity to better serve certain clients and purposes. Fine-tuning of the systems is a continuous process and as we write, new developments occur. However, we believe that as countries are increasingly expressing a demand for support in establishing M&E systems, it is important to recognize how particular arrangements shape and reflect better certain needs and contexts than others, understanding the trade-offs involved. The main conclusion that we derive is that an explicit and thoughtful process of assessing the needs, the focus and the emphasis of the system should serve officials and champions to identify adequate arrangements for the particular context and understand how to better respond to the forces pushing for the creation of new M&E units and bodies.

VII. References

- Briceño, B. And M. M. Gaarder (2009). Institutionalising evaluation: A review of international experience. 3ie-DFID report.
- CLAD-WB (2008). Fortalecimiento de los sistemas de monitoreo y evaluación (M&E) en América Latina y el Caribe, a través del aprendizaje Sur-Sur y del intercambio de conocimientos.
- CONEVAL (2007). Normatividad para la Evaluación de Programas Federales.
- CONEVAL (2008). Informe de evaluación de la Política de Desarrollo Social en México.
- Diario Oficial. México. (2004a) Ley General de Desarrollo Social.
- Diario Oficial. México. (2004b). Lineamientos Generales para la Evaluación de los Programas federales de la Administración Pública Federal.
- Diario Oficial. México. (2005). Secretaria de Desarrollo Social. Decreto por el que se regula el consejo Nacional de Evaluación de la Política de Desarrollo Social.
- Dipres (2008a). Informe de Finanzas Públicas. Proyecto de Ley de Presupuestos del Sector Público para el año 2009.
- Dipres (2008b). System of Management Control and Results-Based Budgeting. The Chilean Experience.
- Dipres (2008c). Statement of the International Advisory Panel for Evaluation and Management Control System.
- EES (2008). Evaluation Connections. The EES Newsletter. August 2008.
- Independent Advisory Committee for Development Impact (IADCI) (2008). *Evaluation independence at DFID. An independent assessment prepared for IADCI by Robert Picciotto.*
- Independent Evaluation Group (IEG) (2007). A diagnosis of Colombia's National M&E System, SINERGIA. ECD Working Paper series, no. 17.
- International Federation of Accountants (IFAC) (2010). Handbook of International Quality Control, Auditing, Review, Other Assurance, and Related Services Pronouncements.
- Kim, John and Nowook Park (2007). Performance Budgeting in Korea. *OECD Journal on Budgeting*, Vol. 7, No. 4, Paris.
- Mackay, Keith (2007). How to Build M&E systems to support better government. World Bank.
- OECD (2002). Glossary of Key Terms in Evaluation and Results Based Management, <http://www.oecd.org/dataoecd/29/21/2754804.pdf>
- Park, Nowook (2008). Does More Information Improve Budget Allocation? Evidence and Lessons from Performance-Oriented Budgeting in Korea. Working Paper to be presented at the Congress of International Institute of Public Finance, August, Maastricht, Netherlands.

Rios, Salvador (2007). CLAD-WB. Fortalecimiento de los sistemas de monitoreo y evaluación (M&E) en América Latina. Diagnóstico de los sistemas de monitoreo y evaluación en Chile.

Rojas et al. (2005). Chile: Análisis del programa de evaluación del gasto público. World Bank.