DFID Department for International Development

3ie International Initiative for Impact Evaluation

Institutionalizing Evaluation: A review of international experience

Bertha Briceño and Marie M. Gaarder

October 2009

Institutionalizing Evaluation: Review of International Experience

Bertha Briceño, Independent Consultant
Marie M. Gaarder, 3ie Deputy Director

October 2009

This publication has been funded by the UK Department for International Development, although the views do not necessarily reflect official policy.

This review is produced by the International Initiative for Impact Evaluation (3ie). 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 campaign to inform better program and policy design in developing countries.

Cover photographs: World Bank / Curt Carnemark, Mexico 1994

Preface

This background report was commissioned by DFID to provide the Government of India with background information and an analysis of international experiences in setting up institutional structures for the evaluation of government programs. A particular interest in the experiences of the Latin-America region and China was flagged.

The authors would like to acknowledge the substantial contributions to this paper by Howard White, Executive Director of the International Initiative for Impact Evaluation (3ie), in particular for providing the inputs for the South Africa, US and UK cases. In addition, the team gratefully acknowledges the substantial time and valuable insights shared with us by Gonzalo Hernandez-Licona, Head of CONEVAL, and Thania Paolo de la Garza Navarrete, Director of Evaluation at CONEVAL, Mexico, and Heidi Berner, Head of Management Control Division, DIPRES, Ministry of Finance, Chile. We would also like to thank Anna Henttinen, DFID, for valuable comments on an earlier draft.

Contents

Tables and boxes

Table of Acronyms

CCT	Conditional Cash Transfer
CNDS	National Commission for Social Development (*Comisión Nacional de Desarrollo Social*
CONEVAL	National Council for the Evaluation of Social Development Policies (*Consejo Nacional de Evaluación de la Política de Desarrollo Social)* (http://www.CONEVAL.gob.mx)
DFID	Ministry of Social Development (*Secretaría de Desarrollo Social*)
DIPRES	Budget Division, Ministry of Finance (Dirección de Presupuesto; http://www.dipres.cl/572/channel.html)
IADB	Inter-American Development Bank
IE	Impact Evaluation
M&E	Monitoring and Evaluation
PAE	Annual Evaluation Program *(Programa Anual de Evaluación)*
SEDESOL	Ministry of Social Development (*Secretaría de Desarrollo Social)*
SFP	Public Comptroller's Office (*Secretaría de la Función Pública)*
SHCP	Ministry of Finance (*Secretaría de Hacienda y Crédito Público*
SINERGIA	The National System for Evaluation and Management for Results (http://www.dnp.gov.co/PortalWeb/Programas/SINERGIA/tabid/81/Default.aspx)

EXECUTIVE SUMMARY

Policy-makers are experimenting with billion's 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. 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, and the availability of funds rather than on their potential contribution to broader development strategies. For this reason, the institutionalization of quality evaluation is necessary in order to turn it into an optimal tool for policy-making. This report looks at the experiences of institutionalizing government evaluation efforts and considers the lessons learnt for countries starting down that road.

This report compares experiences of institutionalizing government evaluation efforts through a discussion of the three leading models in Latin America – Mexico, Colombia and Chile - the non-centralized system of monitoring and evaluation adopted in South Africa, and the policy-learning approach taken in China. Some developed country and international experiences are also briefly presented. The main lessons learned are as follows:

A successful institutionalised Monitoring and Evaluation (M&E) system may look different in different contexts and cultural environments, nevertheless there are the same trade-offs and considerations to be made. With the main objective of monitoring and evaluating the performance of governmental programs, the oversight body should enjoy a high ***degree of independence***, which ultimately translates into higher external credibility, as is the case of Mexico's CONEVAL. Legitimacy of the evaluation effort can also be attained through the establishment of competitive and open procurement processes for the contracting of external evaluations, as is the case in Colombia. The risk to credibility when the system is located under the executive, such as in Chile, can be mitigated with other provisions such as a commitment to public disclosure.

The gains from being 'outside' of government may come at a cost: as the system becomes separated from internal budget or planning authorities, it may have less power to enforce or exert direct influence over the objects of oversight. In the case of Chile, the evaluation unit is located within the Budget Division of the Ministry of Finance and a formal legal mandate requires evaluation of public programs, lending it strong enforcement capabilities. The risk of low ***enforcement capabilities*** can be addressed in alternative ways, however. Support from Congress, fluid communication and promotion of alliances with central authorities, are common strategies to mitigate weak enforcement of recommendations. A complementary strategy to enforce adoption of recommendations is generating a tradition of utilization as a managerial tool rather than a control tool; this is, generating ownership of evaluation by the program's management.

Among developed countries, there are few examples where a national-level body is responsible for evaluation or overseeing the use of evidence – Spain being an interesting exception with their recent creation of a national body in charge of the evaluation of public policies. Rather, they tend to have such bodies at the sectoral level (for education in the US and health in the UK), as well as national-level research funding bodies, which fund academic research but put some emphasis on policy relevance. What we are observing in the case of the UK and the US, as compared to the LA cases, is hence a further degree of specialization.

A majority of countries, both developed and developing, still focus on outcome monitoring rather than evaluation, and particularly impact evaluation. The example of such a system included in this report is South Africa, where public agencies are mandated to establish monitoring systems.

The Chinese experience in developing and implementing economic reforms, included because of the special interest it holds for the GOI, differs in fundamental aspects from standard

assumptions about policymaking. China has been dubbed the 'learning authoritarian state' because it has been able to carry out discretionary policy experimentation *in advance* of legislation. These experimental methodologies are not impact evaluations; rather they are case-studies. There is currently surprisingly little formal, rigorous evaluation work going on in China jointly with the government.

The conclusions highlight some general factors conducive to establishing an institutionalised evidence-based approach to policy-making, including the existence of a democratic system with a vibrant and vocal opposition; and the existence of an M&E champion to lead the process.

More specific recommendations for countries embarking on institutionalizing evaluation include:

- a clear focus on usage and clarity on a client or set of clients that are to be served, and what their interests are;
- a unique and broad legal mandate;
- impact evaluation immersed into broader M&E systems with complimentary monitoring and evaluation instruments;
- build local technical capacity among relevant Ministry officials, program implementers, and local researchers.
- strengthening of data collection and processing systems in order to ensure high quality of data;
- evaluation as an integral part of the programs since their inception;
- legal support from *Access to Public Information or Transparency Laws* is an important asset to back full public disclosure.

I. INTRODUCTION

> *'Building a Monitoring and Evaluation System is a political task, that requires technical elements'.* Gonzalo Hernández Licona, Head of National Council for the Evaluation of Social Development Policies, Mexico.

Policy-makers are experimenting with billion's 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 roads built. In this context, carefully designed and implemented evaluations (in particular, impact evaluations (IE)) have the potential to save lives and improve people's welfare.

However, to date IEs have tended to be selected based on the availability of data, the interest of researchers and donors, and the availability of funds rather than on their potential contribution to broader development strategies. To use IE to achieve optimal allocation of resources requires knowledge of the impact of alternative policies; otherwise there could be a risk of allocating more resources to a program with proven desired outcomes without due regard and information about the optimal coverage of said program or its opportunity costs[1]. For this reason, a comprehensive review of ongoing programs and interventions is needed, under a structure that has the ability to prioritize and the authority to make it happen. In other words, some form of institutionalization of evaluation is necessary in order to turn it into an optimal tool for policy-making. This report looks at the experiences of institutionalizing government evaluation efforts and considers the lessons learnt for countries starting down that road.

The report starts out by highlighting the three leading models and experiences of national evaluation bodies in Latin-America, starting with the Mexican case where an evaluation framework is applied in a federal system, continuing to Colombia where a national evaluation system involves also non-governmental actors, and continuing with Chile where evaluation is used as one of the allocative efficiency tools of the Ministry of Finance. The report then describes the non-centralized system of monitoring and evaluation adopted in South Africa, and the policy-learning approach taken in China, which has been labelled as the 'learning authoritarian state'. Some other institutionalization experiences are also referred to for reference.

For the Latin-American cases the report describes how management of these evaluation bodies is organised, who has oversight and to whom the bodies are accountable, where funding comes from, how evaluation recommendations are acted upon, and how evaluation findings are disseminated. It also describes the institutional relationships of these bodies with other government organisations and other stakeholders, and gives examples of approaches and activities that the evaluation bodies have used to influence attitudes towards impact evaluations and government policy-making more directly. Furthermore, the report assesses the strengths and weaknesses of the institutional structures for independent evaluation bodies and draws relevant lessons that a country like India may consider when setting up an independent impact evaluation body.

In the case of South-Africa, the report describes the attempt at institutionalization of a monitoring and evaluation system (with the emphasis on the former) without the creation of

[1] The opportunity costs to consider could include (i) the effect of alternative programs on the same outcomes; and (ii) the effectiveness of the budget-item that suffered a cut.

a central institution for such a purpose, whereas in the case of China the constraints and possibilities for institutionalizing IE in such an environment are discussed.

After an introduction in chapter I, the report present the five country cases, from the federal evaluation body in Mexico through to the 'learning authoritarian state' of China, in chapter II. Chapter III discusses the trade-offs in the structural design of a monitoring and evaluation (M&E) system, chapter IV the measures of success of an institutionalized evaluation system and chapter V ends with concluding remarks and some lessons.

II. COUNTRY CASES: FROM FEDERAL EVALUATION BODY TO THE LEARNING AUTHORITARIAN STATE

2.1 The case of Mexico: the National Council for the Evaluation of Social Policy, CONEVAL

In Mexico, the leading evaluation entity is the National Council for the Evaluation of Social Development Policies (*Consejo Nacional de Evaluación de la Política de Desarrollo Social – CONEVAL;* http://www.coneval.gob.mx), created in 2004 as part of the Social Development Law. CONEVAL was established with a twofold mission: to measure poverty (national, state and municipal level) and to evaluate 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 -in some cases planning or budgeting units, in other cases special evaluation units- carry out evaluation activities at various degrees, under the guidance and coordination of CONEVAL. Some agencies are more active than others; the Ministry of Social Development (*Secretaría de Desarrollo Social -- SEDESOL*) has been particularly dynamic and has a special unit denominated Social Programs Monitoring and Evaluation Department (*Dirección General de Evaluación y Monitoreo de los Programas Sociales)*. The fact 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 gives the institution a powerful mandate.

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 inspired in theories of new management performance and principles of transparency. 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 institutionalization 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 institutionalized evaluations (impact evaluation and other, such as process evaluation) in the non-social sectors.

2.1.1 Inception

The conjunction of various simultaneous factors in early 2000s cleared the way for the institutionalization of evaluation in Mexico. These factors included: an increasing demand for evaluation from multilaterals, accompanied by technical assistance and support, in particular from the Inter-American Development Bank (IADB); a textbook design of a program for poverty alleviation, Oportunidades (see box 1), in which monitoring and evaluation activities were incorporated in the design by certain evaluation champions (Santiago Levy[2] was a prominent champion); and important institutional changes at the national level in 2000, when a Congress dominated by opposition for the first time after many years demanded close auditing of resources before the upcoming elections.

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 labor 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 behavior and unhealthy life-styles; and (iv) adjustment of the food supplement composition to include a type of iron that would more easily be absorbed.

[2] Mexican economist Santiago Levy is currently the Vice President for Sector and Knowledge at the Inter-American Development Bank. From 1994 to 2000, Levy served as the Deputy Minister at the Ministry of Finance and Public Credit of Mexico, becoming the main architect of the renowned *Progresa-Oportunidades* program that benefits the poor.

Among these enabling factors, the possibly single most important influential factor in the creation of CONEVAL was the strong political pressures from the opposition. The 2001 Budget Law[3] established that all federal programs subject to operational rules should be subject to an annual external evaluation. By 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. Consequently, there was a movement of the opposition within Congress to support the enactment of a Law enabling the creation of a body outside the government devoted to two major functions (i) measuring poverty; and (ii) evaluating social programs. This ended with the enactment of the 2004 Social Development General Law, by which the evaluation process was institutionalized.

In 2005, the newly created CONEVAL found that despite the large amounts of public resources that Mexico had spent in social programs since the 70s, little was clear as to what the returns to those social investments had been. There was an emphasis on outputs rather than outcomes measurement, there was confusion between evaluation and auditing activities, and the social policy decisions were not based on results. The underlying problems that policies tried to address were rarely clear and the prior analysis was missing or poor.

With this diagnosis in mind, challenges to set up evaluation in systemic ways comprised institutional aspects (how to change the rules of the game, and create incentives to carry out and use impact evaluations), technical aspects (capacity and quality of information), and managerial aspects (lack of logical frameworks and problems in identifying what constitutes a program).

2.1.2 Structure and Organization

CONEVAL is headed by an Executive Secretary, and organized in four divisions; two technical divisions, one for Poverty Analysis and the other for Evaluation, and two administrative divisions, administration and coordination (Annex 1). CONEVAL is financed through a direct budget line in the National Budget.

CONEVAL organizes its evaluation work-plan through two separate but interlinked activities; the prioritization of federal programs to be evaluated, and the prioritization of policies. An Annual Evaluation Program (*Programa Anual de Evaluación, PAE*) is defined jointly by CONEVAL, the Ministry of Finance (SHCP), and the Public Comptroller's Office (SFP). The PAE was introduced for the first time in 2007 as a planning tool and has been formalized in the General Guidelines for the Evaluation of Federal Programs, published between CONEVAL, MHCP, and SFP with the main objective of aligning the incentives of the previous evaluations and monitoring regulations. The type of evaluation instruments covered by the PAE are Consistency and Results Evaluations which are required by all federal programs (i.e. evaluation of the consistency of the logframe and monitoring of results), specific evaluations (e.g. evaluation of quality of services of the *Oportunidades* program), and impact evaluations. There are approximately 130 federal programs under the mandate of CONEVAL, 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 (which is an equivalent of 11% of the programs under its mandate), of which some are impact evaluations (approximately 20%), some are specific evaluations, and some are logframe-type evaluations.

A more recent addition to the evaluation efforts is the development of an evaluation-agenda for policy priorities, defined directly by CONEVAL's Board and the executive secretary. Although guidelines for policy evaluations are currently under development, this is found to be a more complicated area with less international practice to draw lessons from. However, the approach taken to policy evaluations seems to be that of identifying priority programs/projects to be evaluated within a certain policy area, as opposed to evaluation of the whole policy.

[3] Presupuesto de Egresos de 2001.

2.1.3 Governance and Accountability

According to the 2004 Social Development General Law, CONEVAL belongs to the executive branch and has technical and managerial autonomy. It is governed by an executive board of six independent academics, the Minister of Social Development, and the Executive Director of CONEVAL. CONEVAL's board of six academics is appointed by the National Commission for Social Development (*Comisión Nacional de Desarrollo Social, CNDS*), 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.[4] Identification of candidates for the six positions is managed through a public bidding process with certain requirements.[5] The six commissioners are appointed for a period of four years, and half of them can be re-elected. The head of CONEVAL is appointed by the executive.

Indirectly, through the legal framework for evaluation, CONEVAL reports to Congress and the general public. The Federal Budget law for 2008 requires the implementation of the Performance Evaluation Framework (*Sistema de Evaluación del Desempeño, SED*), an overall performance system under the authority of the Ministry of Budget, and the SED in turn stipulates the issuance of the PAE that includes the budgetary programs to be evaluated and the types of evaluations to be carried out, and establishes that the information generated by the SED will be disclosed to the public. Also, since March 2008, Congress is receiving from the Executive all the Consistency and Results Evaluations.

The weakest link to date in terms of accountability is the implementation of the recommendations that ensue from the evaluation efforts. The recently emerging practice is that the officials who manage the evaluated programs get to select the 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.

2.1.4 Dissemination

CONEVAL's *General Guidelines* prescribe the dissemination of all evaluations documents and results through the internet websites of the relevant department or entity within 10 business days of the unit in charge of M&E having received said documents. The Guidelines also set out that the information published for each evaluation should contain: strategic objectives, complete text, executive summary and the corresponding appendices, notice of which is the most recent evaluation, and principal results of the evaluation. If applicable, they should also disseminate the current operational rules for the programs, and the agreement of commitments for improving performance.

In addition, they mandate for internet disclosure of contact information of the external evaluator and the program responsible, the type of evaluation, databases, collection data instruments such as questionnaires, formats of interviews, 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, weaknesses, strengths, opportunities and threats, and the recommendations of the external evaluator; and finally, the total cost of the external evaluation, including the source of funding. Mexico also has a general law of access to public information since 2002.

[4] 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.

[5] Criteria for members included: to be or having been members of the national system of researchers; having broad expertise in the subject of evaluation or poverty measurement, and that currently collaborate in tertiary education and research institutions subscribed to Excellence Census of the National Science and Technology Council.

Although CONEVAL has a strong internet dissemination approach, it could benefit from more targeted dissemination events along the line of what is done in Colombia (see section 2.4), involving first technical staff, then managers and heads of units of the program under evaluation, and finally the heads of the agencies, the respective Minister and the Ministry of Finance budget director.

2.1.5 Evaluation Approaches and Guidelines

An example of best practices in evaluation guidelines was recently provided by CONEVAL. 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 programs. This constitutes an unprecedented set of comprehensive and mandatory guidelines for the regulation of evaluation of federal programs.

The purpose of the *Guidelines* is to help regulate the evaluation of federal programs, the preparation of matrices of indicators and monitoring systems, as well as the preparation of strategic objectives of federal public administration dependencies and entities. They are mandatory for federal public administration dependencies and entities that are responsible of federal programs. The guidelines include some general definitions and terminology; evaluation is defined as the objective and systematic analysis of federal programs with the aim of assessing its pertinence and the achievement of its objectives and targets, as well as its efficiency, efficacy, quality, results, impact and sustainability. [6]

The *Guidelines* establish principles and requirements for the different components of the M&E system, including parameters for the establishment of strategic objectives and indicators of the federal entities, indicators matrices, and evaluation. In terms of evaluation, there are provisions that define the type of evaluations, the annual plan of evaluation, and chapters for evaluations of new programs, follow-up of findings and recommendations and dissemination of evaluations and its results.

The types of evaluations are: consistency and results evaluation, indicators evaluation, process evaluation, impact evaluation, specific evaluation (none of the other types), strategic (applied to a set of programs).

Regarding impact evaluations, the guidelines are brief. They define them as the evaluation that identifies with rigorous methodologies the change in indicators at the results level that are attributed to the execution of the federal program, and indicate that the methodologies and terms of reference shall be revised and approved by the SHCP, the SFP and CONEVAL, prior to the contracting of external evaluators. In the future, the Guidelines could benefit from including best practice examples of Terms of Reference for the 4 most common approaches to IE (Randomized Controlled Trial, Propensity Score Matching, Regression Discontinuity, and Instrumental Variable).

2.1.6 Use of Evaluation Findings

CONEVAL, jointly with the Ministry of Finance (SHCP) and Public Comptroller's Office (SFP), published in late 2008 the document *'Mechanism for the follow-up of aspects susceptible to improvements identified in reports and external evaluations of Federal programs'*, aimed at: establishing a general procedure to track improvement aspects derived from consistency and results evaluations and design evaluations; defining people in charge of setting up instruments for such follow-up; and, articulating evaluation results with the performance system managed by the SHCP. A technological system for this was also designed.

The follow-up process involves four steps: analysis, classification and prioritization, instrument formulation (definition of commitments, activities and time span to solve the problems), and results dissemination. Aspects to improve are classified into three types

[6] Lineamientos generales para la evaluación de los programas federales de la administración pública federal. Diario oficial, 30 de marzo de 2007.

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.

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. Some examples of the recommendations identified upon the results of this exercise were: the promotion of mechanisms for identification of potential beneficiaries and target population of the federal programs, improve the design of the social programs, improve effective coordination among institutions and programs, improve information systems of social federal programs, some particular recommendations for the education and health sectors, and some recommendations on measuring results and coverage.[7]

The assessment of the initial implementation of CONEVAL's General Guidelines carried out recently by the WB, lists among its recommendations increasing the use of evaluations results and strengthening its communication strategies, implying there is room for improvement in these areas.

2.2 The case of Colombia: the National System for Evaluation and Management for Results, SINERGIA

In 1994, Colombia established SINERGIA (http://www.dnp.gov.co/PortalWeb/Programas/SINERGIA/tabid/81/Default.aspx), 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 centers, private firms and consultants), governmental agencies, plans, policies, and programs (as objects of M&E, recipients and users) and other producer and recipients of M&E information (statistical institutes, civil society organizations, congress, media).

Thus, 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 through seminars and training events. The normative framework also provides for different units of the agencies and ministries to carry out regular M&E activities, in particular planning or budget units. In exceptional cases, some special evaluation units have been established.

However, involvement in M&E practices varies considerably across different government agencies and ministries, depending on how strongly the M&E culture has permeated the organization, and the particular interest of the agency's head. Accion social, the Presidency's arm for implementation of social programs, has been particularly active. Other M&E systems recently developed or in development by other units within NPD include a system for monitoring of recommendations of the Government policy documents (SisConpes), and the monitoring system for execution of investment resources and products associated linked to the bank of projects (Suifp).

[7] Informe de evaluación de la Política de Desarrollo Social en México, CONEVAL, 2008.

DEPP, as leader and coordinator of M&E activities, is generally recognized as the agency with the technical expertise to support the various agencies in their IE endeavours. It provides support in the construction of TOR, 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.

DEPP is supported by a much weaker legal mandate than CONEVAL (but covering all sectors rather than just social), relying to a large extent on its technical expertise to attract institutions and on the interest of each program/institution head in carrying out evaluations. This approach will tend to favour 'stronger' programs and institutions, leaving perhaps those most in need of evaluation the possibility to opt out. Nevertheless, it is expected that the latter is not a long-term sustainable attitude.

2.2.1 Inception

Colombia started building its current M&E institutions in the early 90s. It can be said that various factors have contributed to the institutionalization of evaluation. First, the construction of the M&E system in Colombia was related to a broader historical process that finalized with the 1991 constitution, by which the country signed a new social agreement that placed a large emphasis on the participatory character of the democracy and on the role of social control in society. The 1991 constitution, and later on, Law 152 of 1994, explicitly assigned to NPD the mandate for promoting evaluation and performance-based management in the public sector. Since then, the government introduced a series of laws, decrees and regulations to further support the system and its instruments.[8]

A decree in December of 1992 created the special division for evaluation and management control within NPD, and conceptualization of the system began in the mid-90s. It initially comprised two modules: internal performance evaluations (i.e. not carried out by an independent agency) and external evaluation of strategic topics.

A second factor was that, after the experience with the evaluation of the Mexican conditional cash transfer program, Progresa, the multilaterals were fostering evaluation of social programs. Overall since the late 90s, there had been more demand from donors for evidence on whether development projects work. And in this sense, rigorous techniques for evaluation, that previously were mainly confined to academic circles, came to play an important role for those involved in development work.

In 2000, a social safety net was launched to offset the effects of the late 90s economic crisis. The so-called Red de Apoyo Social (RAS), included three social programs that were early on identified by multilaterals as promising projects to be evaluated. Thus, they supported earmarking of a 1% of principal in the RAS loan documents, to carry out independent evaluations. To fulfil this objective, a group was formed within the agency that executed the RAS programs. This group was then integrated into the special evaluation division of NPD, and was the basis for what today constitutes the strategic evaluations group of DEPP that commissions and manages evaluations of major government programs.

Another important factor that allowed resurgence of the system after a stagnating period during the late 1990s was the endorsement that President Uribe's first administration gave to the management for results culture, and the relevance he gave to the monitoring system of goals and to monitoring information, in general.

[8] Two preceding developments during the 80s were the creation of the Bank of national investment projects in 1989, in charge of ex-ante evaluation, and the system for multilateral projects evaluation and monitoring –SISEP, established to support management of IDB and WB loans. Related regulations incorporated along the years included decree 2167 in 1992, Conpes 2688 in 1994, resolution 063 in 1994, Law 152 in 1994, Conpes 2790 in 1995, Conpes 2917 in 1997, Conpes 3100 in 1999, decree 1363 in 2000, Conpes 3106 in 2001, Conpes 3117 in 2001, Conpes 3294 in 2004, Art 132 in Law 1151, 2007and Conpes 3515 in 2008.

The challenges to set up evaluation in systemic ways in Colombia have essentially been similar to those faced by CONEVAL in Mexico, including how to increase demand for and use of evaluation, as well as how to increase the supply-side of the evaluations market. DEPP may have an edge over CONEVAL in terms of reinforcing the supply-side, in that it allows for greater competition by organizing competitive biddings for evaluations, and for private organizations to participate in these.

2.2.2 Structure and Organization

DEPP is the technical secretariat of SINERGIA. It is one of the 10 technical directorates within NPD. DEPP is headed by a technical director, and comprises approximately 24 technicians at the professional level and 6 support staff, including a specialized lawyer, administrative, financial and contractual assistants and two secretaries.[9]

Staff is organized under four units: two large ones, the monitoring for results and strategic evaluations units; and two small ones, the dissemination and accountability unit and the performance for results unit (See Annex 2).

DEPP consultancy staff and dissemination activities are financed mainly through NPD's investment budget, on a rolling basis. Resources for major evaluations come primarily from the programs; some evaluations have had support from multilaterals which helped to earmark resources for evaluation within the loan budgets. Other resources for evaluation have been incorporated in certain social loans that included a special line for evaluations, to be co-executed by DEPP.

The evaluation agenda is revised and approved by an Inter-Sectoral Evaluation Committee (IEC), chaired by NPD's deputy director and including representatives from the Ministry of Finance, NPD directorates, and principal sector ministries, with the role of coordinating evaluation processes, approving priorities of programs to be evaluated, approving methodologies, and considering the results that may contribute to improving the formulation of policies. It has functioned on an ad-hoc basis since 2002, but a provision under the development plan Law in 2007 formally mentions the creation of this committee so as to provide for its further legal development. Conpes 3515 of 2008 establishes that all central ministries and agencies should inform the IEC of any impact evaluation planned in order to get feedback and recommendations on the design of the evaluation.

2.2.3 Governance and Accountability

DEPP is headed by a technical director, responding directly to NPD's deputy director and a general director, who have the status of Minister and Vice-minister, respectively.

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

By being located within the Ministry of Planning, DEPP loses some of its claim to autonomy (compared to CONEVAL), without gaining direct influence over budgeting decisions (compared to DIPRES). This position could have been partly remedied by clear public disclosure laws (see below) but this is also currently lacking.

2.2.4 Dissemination

SINERGIA carries out an intensive dissemination process within the government and program stakeholders for each evaluation. This comprises a first stage of revision and discussion with the technical staff from the different units involved and DEPP's evaluation group, a second

[9] DEPP staff work on a consultancy basis, they have not been incorporated permanently within DNP; this has been considered as an obstacle to Sinergia's full institutionalization. This is also partly due to legal provisions for the rationalization of public servants that came into place as part of the late 90s economic crisis.

presentation with the managers and heads of units of the program under evaluation, DEPP and the sector directorate of DNP, emphasizing findings and recommendations, and a third presentation stage with the heads of the agencies, the respective Minister, the Ministry of Finance budget director, the Counsellor Ministry, the General Director of DNP, and staff. After this, it is intended that an improvement plan is agreed between the programs and DEPP, and that compliance of commitments is followed-up by DEPP. The latter is currently in implementation and a technological system for monitoring of compliance with commitments derived from recommendations is currently under development.

Regarding the partial and final reports of the firms, they are subject to DNP's overall procedures on publications, by which a Committee revises and approves before public disclosure, as well as DNP's Director and Sub-director. Regular practice of DEPP has been to publish summary reports on the internet website as well as the databases for public use.

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 including academics, civil society members and stakeholders. Notwithstanding these efforts, full disclosure of original reports by consultants is still needed, and the country still lacks a law of access to public information. It can be said that location within the government somehow limits the full disclosure ability of the system.

2.2.5 Evaluation Guidelines

SINERGIA follows de facto general principles[10] for the commissioning of evaluations, including the following: external contracting, standard bidding processes, independence, objectivity, and methodological rigor according to the nature of the program under evaluation. The evaluation group aims at ensuring adequate involvement and collaboration from the program's management while at the same time safeguarding free and objective assessment of the firm during all stages of evaluation, including the terms of reference preparation, data/information collection, and discussion and dissemination of results.

The main task of the evaluation group is to provide technical oversight of all evaluation products and facilitate evaluations while they are being undertaken.[11] All evaluations comprise at least the following products: a methodological report, a report on instruments for data collection, a field work report, and a complete final report with analysis of evaluation results, and a summary report by DEPP. As part of the quality control mechanisms, it has been introduced recently a practice of external peer reviewers of the main evaluation products. The external peer reviewers are prestigious international experts asked to prepare a referee report to feedback the consulting firm on the final report before closing the evaluation.

Over the years, SINERGIA has established a classification of evaluations undertaken, including impact, results, operational, executive, and institutional. *Impact* evaluation is defined as allowing identification of changes generated by an intervention on the final beneficiary. It is the most demanding type of evaluation since it requires construction of treatment and control groups, and collection of baselines. *Results* evaluation is defined as the analysis of effects on the final beneficiary based on comparisons at different moments of time, without counterfactual group (before-after). *Operations* evaluation is a rigorous analysis of macro and micro processes of an intervention aimed at making recommendations on the program's organizational dynamics. *Executive* evaluation is a detailed analysis of a program structure, in terms of its design and implementation, and based on a standardized

[10] These are not legal, nor mandatory, but de facto principles of work within the DEPP evaluations group.

[11] This includes the following: preparing and concerting the terms of reference, revising the proposals, arranging and participating in the committee that grades the proposals, preparing the grading report, participating in the negotiation with the firm selected, facilitating the contracting arrangements, leading and facilitating all meetings during the evaluation course, monitoring the execution schedule, revising the instruments for data collection, consolidating comments to the partial and final reports, coordinating the adjustment plan and monitoring implementation of recommendations.

questionnaire. It allows a qualification of the program in the different categories. Finally, the *institutional* evaluation is an analysis of a program based on the institutional arrangements in which it operates. It is used to measure the effects of structural reforms on programs or institutions.

No further principles or legal statements regarding evaluation or impact evaluation have been produced by the system, but these are intended to be developed within the new World Bank credit operation that began implementation during 2009.

The development of detailed Guidelines along the lines of those developed by CONEVAL would allow a wider range of programs to get access to technical assistance, at least in an indirect fashion. DEPPs current involvement in the evaluation contracting processes, in quality supervision, and in the dissemination activities limits its ability to cover many programs.

2.2.6 Use of Evaluation Findings

SINERGIA's focus has been on the utilization of evaluation information by the program managers, given its collaborative nature, demand-driven orientation and limited enforcement powers. Monitoring information is used extensively by the President and his office as a control tool. The downside is the limited use from budget authorities and Congress. Recently, the system has engaged in the construction of a system to track commitments from recommendations derived from each evaluation, including executive and impact types. There is no enforcement for adoption of recommendations but they are generally implemented because of high ownership of program 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. Some examples are included in table 1.

Table 1. SINERGIA. Some examples of changes derived from evaluations findings

Familias en Acción Evaluation (CCT program)	▪ Removing non-eligibility of municipalities without banking ▪ Removing non-eligibility of Beneficiary Community Households (HCB) beneficiaries ▪ Reducing subsidy amount for primary education (urban) ▪ Introducing gradually increased amounts for secondary subsidies in 2 different schemes (urban) ▪ Introducing prizes for secondary graduation (urban)
Urban Housing Subsidy	▪ Reforms to the new Urban Social Housing scheme (VISU), included in national development plan (2006-10). ▪ New Social Housing loan (VIS) operation (2008-2011) ▪ Separation of administrative activities between Fonvivienda and Ministry (art. 138 de la Law 1151 de 2007) ▪ Quality improvement for subsidized housing ▪ Revision and strengthening of the auditing scheme ▪ Further outsourcing of the subsidy fund
Non-formal Musical training program (Batuta)	▪ The crowding out of non-displaced children was submitted to the Board Committee for revision ▪ Campaign for donor fund-raising is supported in the positive evaluations results. ▪ Design of a system to monitor children and instructors' evaluations to measure quality of service. ▪ Design of a communications campaign to children and families covering issues of duration and characteristics of the program.

2.3 The case of Chile: the Management Control Division at DIPRES

In Chile, the management control division within the budget department of the Ministry of Finance, Dipres (http://www.dipres.cl/572/channel.html), is the unit that leads the system for evaluation and management control, under which the evaluation of programs and institutions is framed. 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.[12]

Evaluation of programs and institutions is one among the four areas of work that have been developed and introduced especially since the early 90s. The remaining three comprise: monitoring and supervision instruments (performance indicators since 1993, comprehensive management reports since 1997, strategic definitions since 2001, standard presentation of programs to budget since 2000); institutional wage incentive mechanisms (management improvement program since 1998 and refined in 2001 and 2003, incentive to physicians since 2003, and institutional efficiency goals since 2007); and the public management modernization fund since 2008 (see http://www.dipres.cl/572/propertyvalue-2131.html).

Within the evaluation of programs and institutions work area, the instruments introduced for evaluation of *programs* are: the governmental program evaluations, since 1997, the impact evaluation of programs since 2001, and the evaluation of new programs started in 2009. Regarding evaluation of *institutions*, there is one instrument introduced in 2002, the comprehensive evaluation of spending, comprising features of organizational design, consistency with strategic definitions, organizational and results aspects.

The broader picture of government M&E activities includes in addition the system for monitoring governmental programming, known before as the system of presidential goals, dating back to the early 90s, and located under the Secretary of the Presidency (Segpres). The main purpose of this system is the monitoring of the annual programmatic agenda of the government, under which matrices of commitments of ministries are established.

2.3.1 Inception

The origins of the system date back to the early 1990s, with an initial period of consolidation of public reforms. In 1993 a pilot plan of modernization and the Inter-ministerial Committee of Modernization developed the first instrument; performance indicators. Especially since 1994, during the Frei administration, a sequence of instruments was introduced. As from 2000, the administration of President Lagos promoted a more integrated vision of state modernization, with a revision and consolidation of the instruments and the creation of the management control division to implement the evaluation and management control system.[13]

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 instruments underwent further revisions and strengthening during the early 2000s, and additional instruments were incorporated.[14]

The program of evaluation was launched in 1997, after a protocol of agreement with Congress in 1996, aiming at strengthening information provided for budgetary decisions. The program of evaluation responded to a particular demand from the Legislative, seeking further quality information and influence over decision-making. The first evaluations introduced in 1997 were only of the rapid or desk-review type, based primarily on secondary sources. As from 2001 the government introduced impact and in depth evaluations. (World Bank, 2005: 1; Dipres, 2008). In 2003, a formal legal mandate requiring evaluation of public programs was introduced.[15]

Indeed, as the International Advisory Panel for Evaluation and Management Control System stated in 2008, "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".

[12] The objective of the evaluation and management control system is providing performance information and introducing practices to improve the quality of public expenditure improving resource allocation, improving the use of resources, and improving transparency.
[13] World Bank, 2005: 30.
[14] WB-CLAD, 2007: 27
[15] Dipres, IFP 2009: 128

2.3.2 Structure and Organization

The management control division is one of the four divisions and two sub-directorates in which the budget directorate of the Ministry of Finance, Dipres, is organized.

The management control division comprises approximately 32 people organized in three units or departments: the evaluation of public programs unit, the public management unit and the technical assistance unit (introduced in 2008). The public management unit is in charge of the monitoring and follow-up instruments, and the institutional wage incentive mechanisms. The technical assistance unit was created to assist the programs in defining M&E data requirements and ensuring incorporation of them within the programs' information systems, including beneficiaries, baselines and indicators. There is a mix of consultant and permanent-based staff, in accordance with standard public service practices (See Annex 3).

The definition of an evaluation agenda is closely linked to the budgetary annual cycle, and is supported by Congress through the signature of a protocol for selected programs to be evaluated, which occurs in November every year. 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).

2.3.3 Governance and Accountability

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. The Congress has a say in the approval of the protocol of selected programs to be evaluated, it can request the inclusion or removal of certain program or institution within the annual evaluation plan. Seemingly, the Congress has not been very active in modifying the evaluation agenda (Rojas et al. 2005:8).

2.3.4 Dissemination and Institutional Relationships

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.[16]

All information generated by the 3 evaluation lines of Dipres is of public character. This materializes in the distribution of final reports of each evaluation to Congress, and to the public institutions with decision making Powers, and with public availability of reports in Dipres Internet website. Since 2003 evaluation findings are also presented before a special commission of Congress, *Comisión Especial Mixta de Presupuestos del Congreso*. Also, summary or brief reports of evaluations are part of the information that accompanies the Budget Law Project every year, (Dipres, IFP 2009: 129).

2.3.5 Evaluation approaches and Guidelines[17]

The principles for evaluation followed by the division for management control, although not legally formalized, are generally established for all type of evaluations:

- Independence: the evaluation must be external to the responsible institution and ministry. They are carried out by independent evaluators through panels of experts or universities and consulting firms.
- Transparency: the results must be of public character.

[16] *Ley* No. 20285, august 2008.

[17] This section draws extensively on presentations by Dipres. See Berner (2008) and (2008a). It also draws on the International Advisory Panel Statement, Sept. 2008.

- Technically-suited: the evaluation must be pertinent and objective, this is, it should be founded on strictly technical records.
- Timely: the evaluation should provide information in adequate timing so that it supports decision-making processes.
- Efficient: the cost of the evaluation must correspond with the results expected (evaluative judgments).

Also, the evaluations are meant to be budget-related, in the sense that results of the evaluation are taken into account during budget preparation, and they must have a counterpart in the ministries or agencies in charge of programs. Finally, there is definition of commitments to incorporate recommendations from the evaluation and there is follow-up of commitments' compliance.

The evaluation of *programs* line of work includes three instruments. The first is the *governmental program evaluations*, introduced in 1997 and aiming at analyzing consistency of objectives, and organizational, managerial and results aspects at the outputs level (coverage, targeting, among others), and completed between 6 and 12 months. The two other instruments are *impact evaluations*, since 2001, aiming at assessing intermediate and final results on beneficiaries, using control groups and econometric techniques; and the *evaluation of new programs*, introduced in 2009 and aimed at expanding impact evaluations since design. The emphasis of the new programs evaluation instrument is the early design of evaluation for programs at their beginning, as opposed to the earlier evaluations which had an ex-post character and suffered important limitations due to the lack of baselines, information and the impossibility of rigorous designs (Dipres, 2008).[18] Lack of data, design weaknesses and program structural nature were identified as constraints to the quality of evaluations, which was deemed as uneven (Rojas et al. 2005; Mackay, 2007).

The way the new IE instrument has been set up is of particular interest. First, it is stated that the Evaluation of New Programs (EPN) sought to assure that Chile remained in the leading ranks of countries with systematic evaluation processes by updating evaluation procedures and processes to world frontier levels. Thus, the EPN line aims at designing the evaluation at the beginning of each new program; establishing control groups, based on randomized trials whenever is possible; and establishing an international advisory committee to periodically review and assess the process of evaluation.

The most relevant feature is that the EPN counts with the technical support from an International Advisory Panel, made up of well renowned international professors in the IE field.[19] Also, there is an alliance with a local research center, the Centro de Microdatos from the University of Chile, which is a leading center in data collection with extensive evaluation experience.

The International Advisory Panel for Evaluation and Management Control System fulfils an important role, giving Dipres recommendations regarding the technical design of evaluations of new programs, as well as regarding the necessary data collection. It will also support the development of the evaluations and the analyses of their results.

The first statement of the Advisory Panel, in 2008, included very detailed principles for the new evaluation line that Chile is implementing. For their relevance, following is the extracted transcript.

[18] This is a possible explanation for the fact that, in contrast with the Colombian or Mexican experiences, the Chilean impact evaluation line of work has had limited external exposure and resonance. Although the line started in 2001, the only well known evaluation is the Chile Solidario IE, led by the World Bank and Mideplan, which has been object of numerous discussions.

[19] Professors Jere Behrman (University of Pennsylvania), Orazio Attanasio (University College of London), Paul Gertler (University of California, Berkeley), Petra Todd (University of Pennsylvania). It includes local participants as well, professors David Bravo and Claudia Martinez, both from University de Chile.

Principles for Evaluation of New Programs, extracted from the International Advisory Panel Statement, Sept. 2008

1. Extending extensively the use of experimental methods, the "gold standard" of evaluation procedures (even if this happens gradually).

2. Initiating the evaluation process much earlier in project development, preferably at the time that new programs are being designed, new innovations of programs are being planned, and expansions of existing program are being planned. Pilot projects using experimental designs should be the norm for new programs or for substantial modifications of existing programs.

3. Utilizing the best methods available – regression discontinuity, propensity score matching, instrumental variable (IV) estimates and structural models – for evaluations for which experimental methods are not possible or are too costly (as has been happening since 2001, though hampered by usually not having good baselines).

4. Enhancing the data base for ongoing evaluation both by increasing new data collection and increasing the links and the facility of using existing administrative and other data.

5. Coordinating in the Budget Office all the evaluations.

6. Assuring that all evaluations are as "arms length" as possible in order to be as objective (as in the current evaluation system in the Budget Office).

7. Assuring that evaluations are as transparent as possible (as in the current evaluation system in the Budget Office).

8. Systematizing further the integration of evaluation into the budgetary process.

9. Establishing an international advisory committee to periodically review and assess the process.

There are also a couple of additional explicit statements:

"Evaluations are based on data, the development and collection of high quality data must be a priority. The possibility of linking different data sources and to link them also to administrative sources is essential for the development of a good evaluation strategy."

"Evaluations should be independent and detached from the institutions that manage the programs. This makes them credible. This is a good reason to centralize all evaluations centrally in the Budget office. This would also help the integration of evaluations in the budgetary process, which in turn provides the right incentives to the production of objective and high quality evaluations."

2.3.6 Use of Evaluation Findings

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.

The 2008 IFP report by Dipres offers enlightening information in this connection: between 2000 and 2008 there have been 174 programs 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).

It is generally accepted 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; however, managerial usage or ownership from the head of 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 (Mackay 2007: 29).

2.4 The Case of South Africa: the Government-Wide Monitoring and Evaluation System

South Africa adopted a Government-Wide Monitoring and Evaluation (GWM&E) system in 2007. This is not a centralised system of M&E. Rather it is a framework for the M&E systems of public agencies which are legally required to establish an M&E system. The framework identifies three 'data domains': Program Performance Information, Statistical Data, and Evaluation. The Treasury is the lead institution for the first of these, having issued the *Framework for Program Performance Information* for 2007. The government statistical office, Statistics South Africa, has the responsibility for statistics and issued the South Africa Statistical Quality Assessment Framework (SASQAF) in 2008 (Statistics South Africa, 2008). The Presidency, which published the framework document for GWM&E (RSA Presidency 2007), is developing a framework for evaluation.

Each public agency is required to have an M&E strategy, which is integrated with the overall agency management system. Various events have taken place, and inter-agency groups formed, to share experiences in developing and implementing these M&E systems. In addition each agency is expected to undertake capacity building activities for both producers and users of the M&E system.

To date the emphasis has been on monitoring, which is seen as a pre-condition for effective evaluation. The GWM&E framework emphasises developing performance indicators which capture the underlying program logic of an agency's activities – hence the framework is forcing public agencies to explicitly lay out the theory underlying their interventions, analogous to CONEVAL's requirement to adopt log frames. Hence there is a focus on outcome monitoring, starting with 72 national-level 'core indicators'. Impact evaluation is defined in the framework document -"impact evaluations examine whether underlying theories and assumptions were valid, what worked, what did not and why" (RSA Presidency, 2007: 2) – but not otherwise mentioned. However, the GWM&E seminar series has included a presentation on an on-going rigorous IE on land reform being carried out with World Bank assistance.

This is not to say that there have not been previous impact evaluations in South Africa. Most notable are evaluations of the 2004 Social Assistance Act which introduced a number of transfer payments, including old-age pensions, a disability grant, and a child support grant (CSG, an unconditional cash transfer to poorer households with children). A number of studies by international researchers have demonstrated the positive effects of pensions and the CSG on poverty and child health and nutrition, which are cited by the responsible government department. Other areas with impact evaluation include microfinance and HIV/AIDS interventions. The number of studies compares favourably to other countries in Africa, reflecting in part a more developed statistical system and greater capacity in the academic system, though most IEs are by, or in collaboration with, international researchers.

In summary, the South African government has taken recent steps to promote M&E. This is not a centralized system, but a framework for the mandatory M&E systems to be adopted by public agencies, with an emphasis on a program theory approach combined with outcome monitoring. There has been no systematic promotion of impact evaluation. However, there are a fair number of studies, reflecting good data availability and collaborative efforts with international researchers.

2.5 The case of China: the state of current evaluation efforts

China has been dubbed the 'learning authoritarian state' (Heilmann, 2008).[20] Indeed, the Chinese experience in developing and implementing economic reforms differs in fundamental aspects from standard assumptions about policymaking. A core principle of policymaking in rule-of-law systems is that administrative implementation must come after parliamentary legislation or executive regulation and must be based on formalized and publicized general rules. In other words, the potential impact of the policies under consideration must be assessed ex ante without being able to test new policies in practice and obtain realistic information about the potential effects.

In China's reform experience, 'many successful innovations have been the result of administrative "groping along," that is *experimentation during implementation..'* (Heilmann, p.3). Frequently there has been little policy impact assessment and administrative coordination prior to the testing of new policies, rather, discretionary administrative experimentation *in advance* of legislation has played and still plays a crucial role in China's policy process.

The Chinese pattern of experimentation focuses on finding innovative policy *instruments*, rather than defining policy *objectives* (the latter remains the prerogative of the top political leadership). The experimental process is open to decentralized initiatives, thus allowing local officials to become initiators and active participants in the reform drives. Nevertheless, once a pilot project is deemed a success (or not), it is the higher-level decision-makers who decide on the consequences for the initiative and possible scale-up. Heilmann suggests that a majority of major economic reform initiatives in post-Mao China were prepared and tried out through pilot projects (also known as "experimental points" or 'model projects') before they were universalized in national regulations, and offers state-owned enterprise (SOE) restructuring and bankruptcy laws as two important examples. Other areas of experimentation have included the household responsibility system, township and- village enterprises, and special economic zones.

The Chinese-style experimentation takes three distinct forms: (1) regulations identified explicitly as experimental (i.e., provisional rules for trial implementation); (2) "experimental points" (i.e., model demonstrations and pilot projects in specific policy areas); and (3) "experimental zones" (specially delineated local jurisdictions with broad discretionary powers to undertake experimentation). Strikingly, no fewer than *half* of all national regulations in China in the early to mid-1980s had explicitly experimental status.

A number of drawbacks and distortions of the Chinese-style policy experimentation have been pointed out, in particular by Chinese academics, including (i) favouring rent seeking behaviour by local officials; and (ii) if initiated as prestige projects of the top policymakers, pilot projects are often not allowed to fail (implying it is being tweaked until it can produce successful results), with detrimental effects when the policy is generalized. Nevertheless, in his concluding remarks, Heilman notes that *'..it is this particular approach to policymaking that has helped to create a learning authoritarian state and that has facilitated policy and institutional adaptation in China's economic reforms'* (Heilmann p. 19).

Enough is known about the experimental methodologies described above to conclude that they were not experimental or quasi-experimental impact evaluations designed to, in a statistically rigorous manner, capture the counterfactual (what would have happened to the statistically-speaking same population *without* the intervention), and would not pass muster by modern standards (Nevertheless, this does not mean that some useful learning was not

[20] 'In 1978, the Chinese Communist Party's 11th Congress broke with its ideology-based view of policy making in favor of a pragmatic approach, which Deng Xiaoping famously dubbed "feeling our way across the river." At its core was the idea that public action should be based on evaluations of experiences with different policies—"the intellectual approach of seeking truth from facts." In looking for facts, a high weight was put on demonstrable success in actual policy experiments on the ground.' (Ravaillon, 2009)

obtained from the case-study approach used). In fact, according to personal communication with academics, as well as a limited survey[21], there is surprisingly little formal, rigorous evaluation work going on in China jointly with the government. Of the 17 programs for which we were able to establish the existence of completed or ongoing evaluations, the demand for the evaluations has mainly stemmed from donor institutions or has been initiated by researchers themselves, with the Ministry of Health being the most notable exception.

This is mainly attributed to the prevalent ***incentives*** in China's government system. Although government agencies today have access to a lot of funds, the combination of pressure to "get the money out into the field" and short rotations (officials often only stay in a post for three years or so) leave little incentive to figure out what works. As has been the case in most other countries until recently, the evaluation effort still centres around inputs and outputs (budget spent, roads built, scholarships issued) rather than outcomes and impacts (education completed, earnings improved). Partly due to the ***history*** of point experimentation with a strong role for local officials, the idea of randomization of project sites is not an easily acceptable approach to Chinese officials. It is considered less risky to carry out the intervention in one's home town or in the jurisdiction of the most capable leader, as this is assumed to minimize the risk of the experiment 'failing'.

In order to move away from these constraints on large-scale evaluations of government programs, upper level leaders (e.g. in the State Council / NDRC / Ministers' offices) would have to come out in favour of making large scale and rigorously designed evaluations part of China's policy implementation plan.

Nevertheless, this is not to say that China's leaders are not interested in the results of well-designed evaluations of innovative projects that might be useful in helping the nation meet its policy targets. One foreign researcher[22] said it is easy to get the cooperation of government agencies to help clear the way bureaucratically for studies to go forward, and people readily volunteer to be on "Implementation Advisory Teams." It is also easy to get government officials to visit projects and provide their inputs, as they are eager to learn and to report on success stories (to their superiors and in more general conferences). The inherent risk inherent is that the potential lessons to be derived from less successful or unsuccessful initiatives may be missed. In part, the form the participation takes can be explained by the fact that officials are protected from failure, since it is the NGO or the research organization that is carrying out the project, while they can take some of the credit for successes. In part, it has to be recognized that many of China's officials today are well-educated academics / professional themselves with a natural inclination for wanting to know whether something works or not.

Hence, the suggested mode to get the authorities interested and involved in IE is as follow: (i) lead with an innovative project that is in the policy spotlight; (ii) get officials interested and active in observing and providing input early in the project; (iii) produce easily understood evaluation results; (iv) allow program officials to take credit and submit reports to their superiors; (v) write policy briefs to top ministerial and national leaders; (vi) disseminate work in the press; and (vii) if possible, work with government officials in scaling up.

2.6 Other Interesting Institutionalisation Experiences

This sub-chapter covers international agencies supporting the use of impact evaluation evidence and national agencies in developed countries in charge of evaluation of national public policies. This is not intended to be a comprehensive review, but rather give examples of existing institutionalization efforts beyond those reviewed for developing countries. Please note that the Spanish, UK and US experiences discussed here do not cover their efforts in evaluating the development effectiveness of their aid (carried out by the Office of Development

[21] A survey was sent to 14 academics who have been involved in the evaluation of Chinese programs.
[22] Personal communication Scott Rozelle, Professor and Senior Fellow, Food Security and the Environment Program, Freeman Spogli Institute, Stanford University.

Planning and Policy *Evaluation* (*DGPOLDE*), the UK Department of International Development (DFID), and the US Agency for International Development (USAID)), but rather of their own national programs.

2.6.1 International agencies supporting use of impact evaluation evidence

(a) The Cochrane and Campbell Collaborations

Cochrane and Campbell (also called C1 and C2) promote evidence-based policy making utilizing systematic reviews of existing studies.[23] Systematic reviews are reviews that follow a strict protocol and usually include statistic meta-analysis. Cochrane, founded in 1993, provides reviews of medical and health-related interventions, the latter including both public health and health management issues. Campbell is a more recent organization, producing reviews in the areas of social welfare, crime and justice, education. The reviews from both organizations cover only RCTs and a subset of quasi-experimental designs. There has to be at least one RCT on a topic for a review to be undertaken.

Neither organization produces primary studies. Both have a similar structure, having subject groups, whose membership is voluntary (i.e. unpaid), comprising academics who produce the reviews and provide the oversight for quality assurance. Non-members can register reviews with the relevant collaboration. The reviews of both organizations are focused on interventions in developed countries. Of the 42 Campbell reviews to date (Cochrane has many more, but these are mostly medical) only one (on school feeding programs) includes interventions from developing countries.

Neither organization explicitly endorses specific policies, nor conducts any advocacy work beyond promoting systematic reviews in the scientific community. The reviews are publicly available for users to act on as they see appropriate.

(b) The International Initiative for Impact Evaluation (3ie)

3ie (*www.3ieimpact.org*) is a new organization who purpose is to enhance development effectiveness through the promotion of using evidence from rigorous impact evaluations. 3ie provides financing for primary studies of socio-economic interventions in low and middle income countries, provides quality assurance services of studies conducted by others, and supports the production of synthetic reviews of existing studies.

2.6.2 The case of Spain: the Spanish National Agency for the Evaluation of Public Policies and Quality of Services

The Spanish National Agency for the Evaluation of Public Policies and Quality of Services (*La Agencia Estatal de Evaluación de las Políticas Públicas y la Calidad de los Servicios, AEVAL, http://www.aeval.es/en*) was created on January 1, 2007. Its main objective is to promote and carry out evaluations of policies and public programs, promote a better use of public funds, improve the quality of public services, and enhance the public accountability of government bodies. Its creation was based on the recommendations of an expert panel of academics, distinguished professionals, and public managers that prepared a detailed analysis of evaluation in Spain with references to international experiences.

The Agency is a public-law body with its own legal personality and with management autonomy, under the Ministry for Public Administration. The Agency evaluates the programs and policies selected each year by the Spanish cabinet, and submits an annual report to Parliament on central Government agencies' efforts to improve the quality of the services

[23] See http://www.cochrane.org/ and http://www.campbellcollaboration.org/.

they provide to the public. A management contract governs the Agency's activities and its relations with the Government, which funds those activities.

To the best of our knowledge, there has been no rigorous impact evaluation (i.e. evaluation that establish unbiased counterfactuals) carried out or overseen by AEVAL to date, but it is included as one of a range of M&E tools that they will use and promote in the future.

2.6.3 The United Kingdom and the United States

These two countries are treated together here on account of the similarities in their systems. Both have national-level research funding bodies (ESRC and NSF respectively), which fund academic research but put some emphasis, particularly ESRC, on policy relevance and user engagement. Neither country has a national-level body responsible for evaluation or overseeing the use of evidence, though the audit bodies (NAO and GAO in the UK and US respectively) carry out work equivalent to process evaluation. But both have such bodies at the sectoral level: for education in the US and health in the UK.

The Economic and Social Research Council (ESRC) in the UK (annual budget just over US$320 million) and National Science Foundation (NSF) in the US (annual budget just over US$6 billion) are responsible for funding research. NSF's brief runs across both natural and social sciences (except medical, which is handled by NIH), whereas ESRC is restricted to social sciences (with councils for natural sciences and medicine, EPSRC and MRC respectively).

Both ESRC and NSF fund 'pure research', however both put some emphasis on what ESRC call 'engaging society'; specifically they 'place the highest importance on the communication of our research findings to policymakers and research users from government, business and finance, the public and voluntary sectors, and the general public'. To this end ESRC produces research summaries of selected research which are similar to policy briefs, but does not engage in advocacy. Since ESRC is the largest source of funding for academic research in the social sciences in the UK, this emphasis on policy relevance has had some effect in shifting research agendas and researchers to be more 'user friendly'. In contrast NSF's emphasis is more focused on keeping the US on the 'leading edge' of scientific development, but the 'broader impacts' of the proposed research are specifically included in the review process.

Both countries have institutions which conduct systematic reviews of evidence in specific sectors. The What Works Clearing House (WCC, http://ies.ed.gov/ncee/wwc/) is part of the Institute of Education Sciences of the U.S. Department of Education. It was set up in 2002 to be 'a central and trusted source of scientific evidence for what works in education', and immediately courted controversy by proposing RCTs as the gold standard and limiting what it viewed as valid scientific evidence to quantitative experimental and quasi-experimental designs. For example, in the most recently posted review on the WWC website at the time of writing, the report notes that 11 studies were found of the intervention, but none of them met WWC criteria so no studies were concluded in the review, so WWC is unable to offer evidence of the program's effectiveness. The reviews are carried out by WCC staff, which is contracted out to a private, for-profit, research organization. WCC also produces standards on conducting reviews. Whilst the site carries disclaimers that it does not endorse any specific policies, it clearly presents itself as providing guidance to educators on which approaches work and which don't.

In the United Kingdom, the National Institute for Health and Clinical Excellence (NICE, http://www.nice.org.uk/) plays a similar role for the health sector. NICE produces guidance for health service providers based on synthetic reviews. NICE guidelines on clinical and cost-effectiveness – which rank treatments by the cost per quality-adjusted life year - can be binding for providers under the National Health Service (NHS), treatments deemed as not being cost effective being ineligible for treatment under the NHS. This role means that NICE is a very powerful organization, but frequently at the centre of unfavourable publicity, e.g. a recent decision to withhold Tyverb, a drug for treating breast cancer (the decision has been appealed by the manufacturer).

Table 2. Summarizing the Systems: Characterization of Evaluation Bodies

Dimension	Mexico	Colombia	Chile
Origins	Political pressure by opposition in Congress, M&E Champion of *Progresa* textbook design, multilateral demand	Constitutional accountability focus, multilateral demand, *Progresa* demonstration effect, President Uribe's championship of management for results administration	Public Reform Program, Congress demand, Budget directors' continued championship
Location	Independent public administration entity	Under the executive; a directorate within the Planning Ministry (DNP)	Under the executive; a division under the Budget Directorate within the MoF
Scope of Mandate	Evaluation of *social* development programs and policies, and measurement of poverty at the national, state and municipal level.	DNP has the mandate to plan, design and organize the systems of evaluation of results and management, for the entire public administration	Improve efficiency in allocation and utilization of public resources assigned to different programs, projects and institutions
Size of evaluation departments and activities	Aprox. 70 people. 119 evaluations during 2007-2008; out of these 106 consistency and results evaluations and 13 design evaluations. Of these, 10 contracted directly by CONEVAL	Aprox. 30 people. Between 2006-2009, 28 evaluations completed, out of these, 9 impact evaluations	Aprox. 32 people. Since 2001, an annual average of 14 governmental program evaluations; and 7 impact evaluations annually
Annual Budget	USD 12.0 million (2008) [24]	Rough estimate for 2009/10: USD 6.7 million (0.003% of GDP) for aprox. 26 evaluations of all types; evaluations finalized during 2007-2009 cost approx. USD 5.3 million[25]	NA
Governance and accountability	Reports to a Board of six independent academics	Reports to DNP's General Director and to Presidency	Reports to Congress, and Finance Minister
Dissemination of Findings	Full disclosure on Internet websites of databases and reports is mandatory by the General Guidelines	Partial disclosure on Internet website of evaluation data and reports, and public discussion seminars. Full disclosure of monitoring information	Full disclosure on Internet website by access to public information Law
Regulatory aspects	The General Guidelines are mandatory principles for the evaluation of all federal programs; Annual program of federal evaluations, PAE; There are Norms including guidelines and models for standardized TOR that federal dependencies and entities must observe	Standardized TOR for the rapid or executive evaluations, not for impact evaluations, which vary according to the programs' nature; No legal or mandatory IE guidelines, ad-hoc principles of quality; Central entities commanded to present to the IEC any impact evaluation planned	Standardized TOR for evaluations;
Scope & enforcement of regulations	Federal programs by federal dependencies and entities; guidelines are mandatory for them; strong legal support	Central entities; limited enforcement capacity, mainly demand-driven by DEPP's technical capacity; very limited regulatory legal support	Central entities; large enforcement capacity based on budget powers and own funding; supply-driven

[24] Information retrieved at the following site: http://www.oracle.com/global/lad/customers/profiles/oracle_snapshot_coneval_2.pdf
[25] Estimate by Bertha Briceño, former Director of Sinergia.

Table 3. Summarizing production, demand and use of evidence

Dimension	Mexico	Colombia	Chile	South Africa	China	Spain	UK	US
Does a centralized IE body exist?	YES	YES	YES	NO	NO	YES	NO	NO
Who finances and produces IEs?	Donors and institution budget; Public Research institutions,	Donors and institution budget; Public and private research institutions,	Govt, researchers	Govt, donors, researchers	Donors; Researchers	Govt	Govt, funding bodies, researchers	Govt, funding bodies, researchers
Who demands IE?	CONEVAL MHCP, SFP, some ministries donors	SINERGIA, (MoF), some ministries, donors	DIPRES/ MOF	Part of GWM&E but focus has been on M not E	Donors, researcher, (Ministry of Health)	Spanish Cabinet	Parliament, Public	Congress, Public
Who uses IE?	MOF, congress, program officials, donors, the public	Program officials, donors	MOF, congress, Program officials,	Little systematic use	Govt, program officials, donors	Parliament	Govt	Govt

III. THE STRUCTURAL DESIGN OF AN M&E SYSTEM: BALANCING TRADE-OFFS

The experiences of the countries analyzed reveal underlying delicate balances in the institutional design of an M&E system. Mackay points out that M&E should not be pursued as an end itself, but its value comes from usage. M&E information is used for multiple purposes: for feeding back into policy and budget decision making and national planning, improving policy analysis and policy development, helping in managerial activities such as program management and staff or institutional management, enhancing transparency and accountability, and many others (Mackay, 2007: 9).

A successful M&E system may look different in different contexts and cultural environments. In general, the development of the systems has not been linear and follows the learning by doing principle. The most crucial aspect that is repeatedly raised by experts as a yardstick of success is the degree of utilization of the information produced by the system. This is also a prerequisite for sustainability.

Notwithstanding the particularities of each country context, it is possible to think that certain structural arrangements are more naturally associated or nurture certain features. In what follows we aim at revising some of these relationships to provide a framework for understanding strengths and weaknesses of the various possible arrangements.

In terms of location, we have observed that some systems are established outside the executive, others are located within the government. Since the M&E systems we analyze have as object of study the monitoring and evaluation of governmental programs, plans, projects, or activities in general, one sensible assumption is that subjectivity increases the closer the M&E unit is to the object of analysis. Common sense also says that an oversight body should enjoy a high **degree of independence** to be able to freely make assessments and fully disclose them, without any improper influence. Therefore, the implication being that out-of-government systems should enjoy a higher degree of independence that ultimately translates into higher **external credibility**. Presumably, the higher the degree of independence, the

better is the reception from clients outside the government such as Congress, media, and civil society (CONEVAL).

A related issue regarding legitimacy is that, despite its location, separation between the evaluator and the evaluated is a well established principle in Latin American systems.[26] The systems promote contracting evaluation externally to consultants, firms, or research centres and tend to have competitive and open procurement processes[27] (Dipres, SINERGIA).

The gains from being 'outside' of government may come at a cost: the downside of the completely external arrangement is that as the system becomes more separated of internal budget or planning authorities, it may have less power to **enforce or exert direct influence** over the objects of oversight. Transparency and accountability utilization might be stronger, at the expense of a lesser utilization as an internal management control tool from the government's center (budget central authority, planning, presidency or internal control office) or as a management tool by the own programs, unless other provisions are in place. Lesser enforcement capability may be present as well when the system is located inside the government, but outside direct budget authorities.

The risk of low enforcement capabilities can be addressed in diverse ways. Support from Congress, fluid communication and promotion of alliances with government central authorities, are common strategies to mitigate weak enforcement of recommendations (CONEVAL with MoF, and SINERGIA with the Presidency). A complementary strategy to enforce adoption of recommendations is generating a tradition of utilization as a managerial tool rather than a control tool; this is, generating **ownership of evaluation** by the program's management, thus inducing utilization –i.e. *voluntary adoption of recommendations* - by the program. The M&E body thus invests highly in demonstrating the benefits of evaluation as a managerial tool, in capacity building activities and in establishing a favourable cultural climate for M&E (SINERGIA and CONEVAL).

Furthermore, these "persuasion" strategies become more crucial in systems where the evaluation agenda is more demand-driven oriented than supply-driven. There is more enforcement capability or a supply-driven evaluation agenda when the system enjoys a strong legal support (CONEVAL) or has a permanent budget line or own financial resources to carry out the evaluations (Dipres), as opposed to when there is weaker legal support, when the programs voluntarily devote resources out of their budgets, are earmarked in loans, and in general, depend more on buy-in to evaluation as a way to improve its own performance (SINERGIA, CONEVAL partially).

Buy-in from program management is also important from the information disclosure viewpoint. The success of M&E activities is largely dependent on the quality and availability of internal information produced by the programs, on their willingness and capacity to generate primary data or recover information on beneficiaries, and on the flexibility to implement pilot changes in the programs or in certain groups. In this sense, it is expected that the higher the ownership of the evaluation, the higher the **insight,** quality and completeness of information provided by the programs for M&E activities. **Quality of input information** can also be sought establishing control mechanisms, external verification or audits and alliances with internal control or auditing offices.

The importance of the **enforcement** capability rests on the fact that utilization of the system as supporting an internal management control function by the government's centre (not necessarily by the program's management) depends on its power to *enforce the necessary adjustments* derived from M&E assessments. 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 the extreme enforcement version, this is complemented with getting support from Congress (Dipres).

[26] As opposed to, for instance, the Office of Management and Budget's approach with Program Assessment Rating Tool, in the USA.
[27] Mexico uses more frequently direct collaboration with research centers.

Location under the budget authority also provides for a higher integration of M&E and the budgeting and executing stages of the public policy cycle.

Other systems are framed under a planning tradition. In such cases, M&E activities are designed to assess implementation progress of governmental development plans or national policy priorities; so that linkages are naturally stronger with the planning stage of the public policy cycle (SINERGIA, CONEVAL).

The presumably lack of autonomy to **disclose** M&E information produced when the system is located under the executive, can be mitigated with other provisions such as a long-standing tradition or commitment to public disclosure, or a particular law or decree (Dipres, CONEVAL). Overreaching laws of access to public information have been recently introduced in various countries and indirectly support the M&E system disclosure ability (Chile, Mexico).

In addition, to the structural arrangements and inbuilt incentives (carrots and sticks) discussed above, which vary from one country to the next, a common challenge has been raised in meetings and/or relevant documentation for all of the developing countries discussed in this report, and that is the ***lack of technical capacity*** among program staff and researchers in evaluation methodologies. Currently, all of the discussed countries rely to a large extent on foreign researchers to lead their evaluation efforts, thereby seriously limiting the supply and in some cases the quality due to the lack of understanding of context. The fear of the methodologies due to lack of exposure to them also limit the demand for more evidence.

Table 4. Elements of Design and Associated Features

	Location			**Financing**		**Clients –Utilization**		
	Inside Government		**Outside**			**Internal**		**External**
	Under Budget Authority	Other/ Planning Authority		Programs	Budget Line	Center: Presidency, MoF	Program Managers	Congress, Media, Civil Soc.
Degree of Independence			v		v			v
External Credibility / Influence			v		v			v
Degree of Enforcement	v				v	v		
Ownership of Management		v	v	v			v	
Internal Insight, quality of information		v		v			v	
Disclosure			v		v			v
Links to budgeting & execution	v				v	v		
Links to planning		v		v			v	

IV. MEASURES OF SUCCESS

As we mentioned earlier, usage of M&E information, including evaluation results, justifies to a great extent investment of resources in impact evaluations and other M&E instruments, and more generally, determines the sustainability of the systems. Idiosyncratic development of each country's system and cultural features shape differently the focus of the system utilization, with various combinations from single to multiple clients and usages. So far, we

have identified internal clients from the Executive, such as Planning and Budget Ministries, and agencies central units, such as planning, budget and, in some cases, special evaluation units, all of which seek to improve the efficiency and effectiveness of resource allocation. Other internal clients include the executing or implementing agencies, which are generally more interested in revising processes, changing and improving managerial practices with a results orientation, 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.

Defining measures of success in terms of utilization is not an easy task, and is an endeavour that the systems only recently are beginning to undertake more carefully. The Independent Evaluation Group of the World Bank (IEG) has contributed with actively promoting some assessments of the systems' performance and diagnoses. CONEVAL recently commissioned an assessment of its General Guidelines for Federal Programs Evaluation from a World Bank team, another team carried out in 2005 a comprehensive analysis of the Chilean public expenditure evaluation program and IEG published a diagnosis of SINERGIA in 2007. The Latin American Centre for Development Administration (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 by studying and analyzing 12 countries, with a standard methodology and a comparative approach, which resulted in a series of individual country studies and a 2008 comparative report. 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; the Chilean World Bank evaluation included a revision of samples of evaluation reports, assessed comparatively against a certain standard criteria, interviews and focal groups. SINERGIA's diagnosis was mainly a case study with in-depth interviews and documentation revision.

Particularly for evaluation, two dimensions have been commonly explored. First, what can be referred to as coverage, a measure of the extent of the evaluation practice in relation to a reference value or universe? Usually, 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 a multiannual agenda or the number or programs in a programmatic classification of the budget.

Second, in terms of utilization of evaluation, it can be said that a sort of incipient consensus is developing towards the follow-up of recommendations, commitments and action plans derived from the evaluations. This can be for instance simpler measures as the number of changes derived from evaluations, number of recommendations adopted; or more demanding ones, as the proportion of the recommendations implemented over the total number of recommendations formulated.

Finally, in terms of final goals of the systems, such as improving quality and efficiency of public expenditure, there have not been clearly established indicators to address such qualities nor to correctly attribute the effect of the M&E systems.

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.[28] In addition, when the system has also an orientation towards influencing budget allocations, further utilization measures could include the change in allocations as a

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

result of evaluation utilization by budget and Congress, or more indirectly, correlation measures with resources allocation changes.[29]

In general, actual assessments of utilization have included so far an idea of coverage, clients' satisfaction surveys, some evidence on adoption of recommendations and commitments, and some anecdotic evidence; mainly in the absence of systematic collection and monitoring of the recommendations and commitments. In any case, the issue is beginning to be addressed in more systematic ways.

V. CONCLUSIONS

This report has analysed and compared several existing institutionalization models, and highlighted the challenges and advantages of each. No unique model for institutionalizing and strengthening the M&E system exists, rather, best practices in different countries depend on the existing technical capacities, the institutional organization (e.g. level of decentralisation, budgeting practice), political context (e.g. role of Congress, credibility), management culture, and leadership of an agency (e.g. Office of the PM, specific Ministry).

Nevertheless, a number of general conclusions and lessons can be drawn from the experiences presented in this report, as well as some more actionable lessons that may prove useful to governments considering setting up an M&E system and institutionalising evaluation.

Some general conclusions and lessons:

- The existence of a democratic system with a vibrant and vocal opposition is an important enabling factor when it comes to the institutionalization of evaluation bodies with some inbuilt independence. Nevertheless, the establishment of such systems are a lengthy process (not yet finished in any of the countries discussed) as democracy also requires extensive information campaigns, consultation processes, and legal and parliamentarian steps (contrary to the 'experiment first' approach that has been used extensively in China).

- The need for an M&E champion to lead the process. A clear powerful stakeholder, such as Congress, the MoF, or the President facilitates triggering the process. The three Latin-American country cases exemplified this.

- The experiences of the countries analyzed reveal underlying delicate balances in the institutional design of an M&E system: location, source of financing, and utilization focus determine to a great extent the trade-offs in desired dimensions such as the degree of independence, external credibility, degree of enforcement, ownership of management, internal insight and quality of information, disclosure ability, linkages to budgeting, execution and planning.

Some more specific recommendations and lessons:

- Utilization is the yardstick for success of an M&E system. 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, the external society, the government's centre (Presidency, budget office, planning office) or program management. Sustainability depends on usage.

- It is important to enjoy a unique and broad legal mandate. The risk of having ambiguities in the legal or regulatory mandates over the agency or unit in charge of the development of the system, or the scope of sectors, is that competing initiatives may appear that undermine consolidation and legitimacy before the agencies and externally. Recent

[29] For an interesting example on this potential measure, examining the correlation between evaluation results and budget growth of evaluated programs in Korea, see Nowook Park (2007, 2009).

developments in Peru, with ambiguous mandates of the Ministry of Finance and the commission for social affairs, and the limited scope of CONEVAL to the social sector exemplify such risks.

- Impact evaluation needs to be immersed into broader M&E systems with complimentary monitoring and evaluation instruments: rapid, process, operations, institutional, and other types of evaluations. The experience seems to be that gradual evolution from less to more sophisticated evaluation instruments is important in helping developing an M&E culture that paves the way for rigorous impact evaluation.

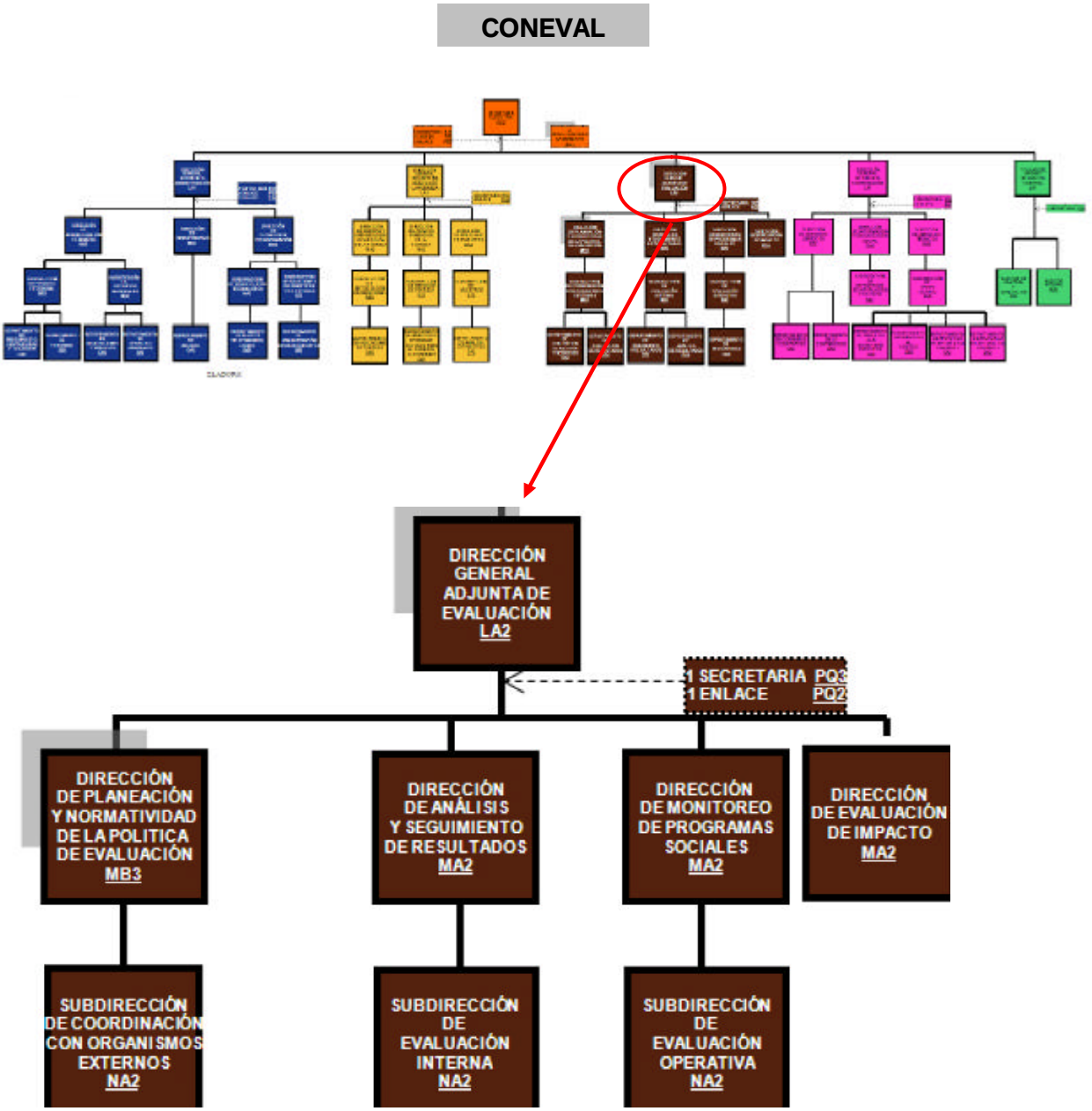
- Fundamental to the production of, demand for and use of evidence/evaluations is the building of local technical capacity among relevant Ministry officials (e.g. in their M&E departments), program implementers, and local researchers.

- Setting up an M&E system takes years, and is likely to be an always ongoing process rather than one that reaches completion. It requires the strengthening of data collection and processing systems in order to ensure high quality of data, the building of capacity, and the willingness to continuously learn from the experience of others.

- Evaluation needs to be an integral part of the programs since their inception. The Chilean experience with development of ex-post impact evaluations and the move towards the new programs evaluation exemplifies this.

- Legal support from *Access to Public Information or Transparency Laws* is an important asset to back full public disclosure, especially in systems located within the executive. The case of SINERGIA's evaluations exemplifies the lack of such.

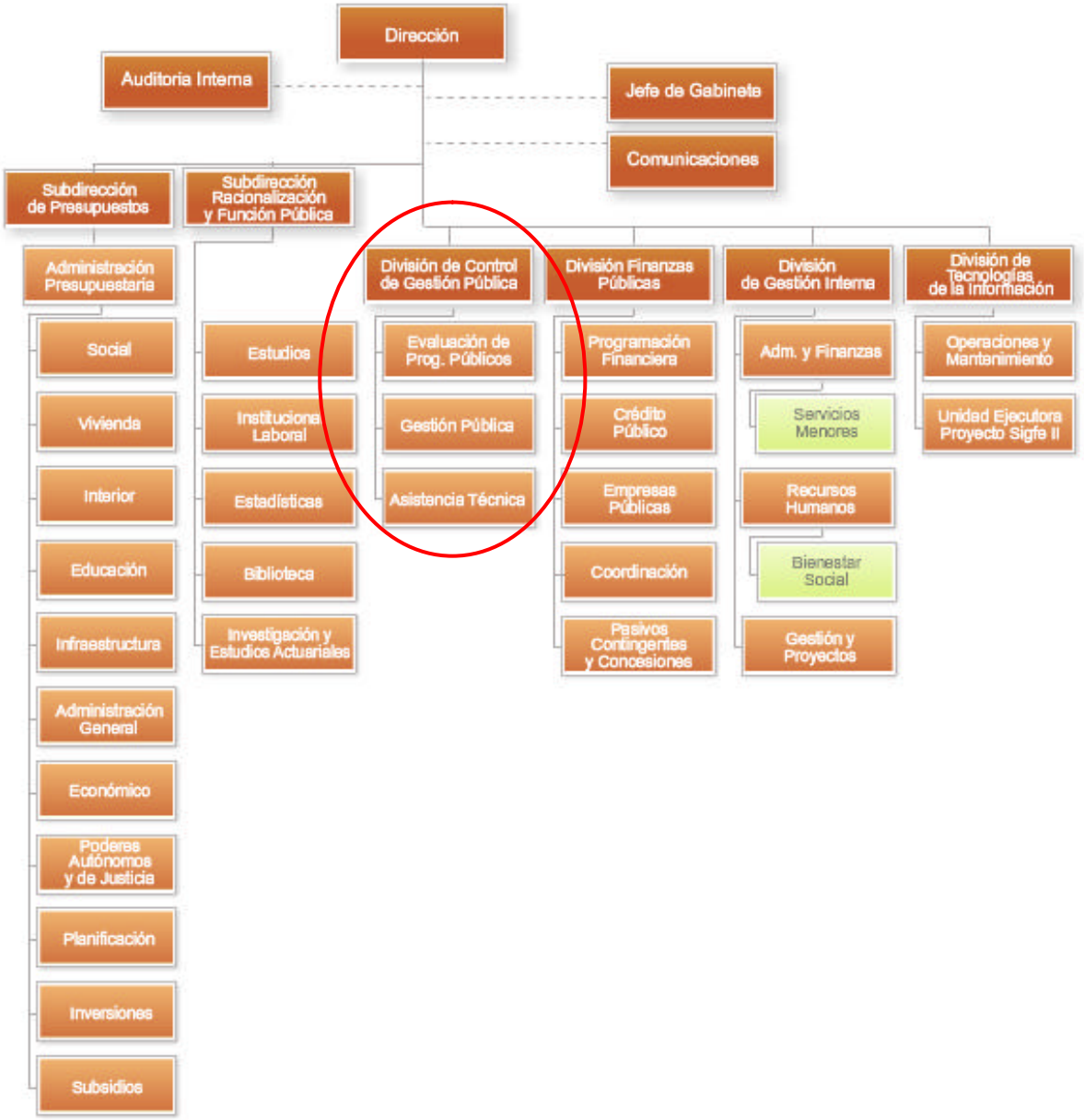
Annex 1. The case of Mexico: Organizational Structure and location of CONEVAL



Annex 2. The case of Colombia: Organizational Structure and location of DEPP

Annex 3. The case of Chile: Organizational Structure and location of the Management Control Division at DIPRES

Budget Directorate (DIPRES) in MHCP



References

Diario Oficial. México. (2004) Ley General de Desarrollo Social.

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.

Diario Oficial. México. (2004). Lineamientos Generales para la Evaluación de los Programas federales de la Administración Pública Federal.

Dipres (2008). Informe de Finanzas Públicas. Proyecto de Ley de Presupuestos del Sector Público para el año 2009.

Dipres (2008). Sistema de Control de Gestión y Presupuestos por Resultados. La Experiencia Chilena - Noviembre de 2008.

Dipres (2008). System of Management Control and Results-Based Budgeting. The Chilean Experience - October 2008

Dipres (2008). Statement of the International Advisory Panel for Evaluation and Management Control System. September 2008

Heilmann, Sebastian (2008); 'Experimentation under Hierarchy: Policy Experiments in the Reorganization of China's State Sector, 1978-2008'; Center for International Development at Harvard University, Working paper No. 172.

Hernandez-Licona, Gonzalo (2009). Impact Evaluations in Mexico. CONEVAL. Ponencia.

Hernandez-Licona, Gonzalo (2009). Construyendo un Sistema de Monitoreo y Evaluación: Un Reto de Política Pública con Elementos Técnicos. CONEVAL. Ponencia.

Mackay, Keith. (2007) How to Build M&E systems to support better government. World Bank.

Medina, Alejandro (2007). CLAD-WB. Fortalecimiento de los sistemas de monitoreo y evaluación (M&E) en América Latina. El Sistema Nacional de Monitoreo y Evaluación de la Gestión Pública en México.

Pérez-Yarahuán, Gabriela (2008). Evaluación de Programas Sociales en México. El caso de la Evaluación de Consistencia y Resultados. CONEVAL. Ponencia.

Ravallion, Martin (2009). Evaluation in the Practice of Development. Lecture at 3ie's Annual Conference (in collaboration with NONIE and AFREA) "Perspectives on Impact Evaluation", Cairo, Egypt.

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.

Notes

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