Closing the Circle for Results-Based Management in Social Policy

Miguel Székely

The idea of introducing results-based management in the social sector has been expanding in the developing world. For this to happen, evidence on the impacts of different programs needs to be available, and the public sector has to use this information by feeding it into the planning, design, and execution stages of policy. This note argues that incentives faced by different actors and the way in which the evaluation function is organized within governments play important roles in evaluation use. The note reviews three types of institutional arrangements for organizing the evaluation function with a primary focus on Latin America. The experiences of Mexico, Peru, and South Africa are used to illustrate that closing the results-based management circle is feasible under any of the three arrangements.

Results-Based Management Cycle

The term "results-based management" has become increasingly popular in social policy circles in recent years. The idea is that social policy decisions will be guided by information on the extent to which different programs improve the well-being of the intended beneficiaries.
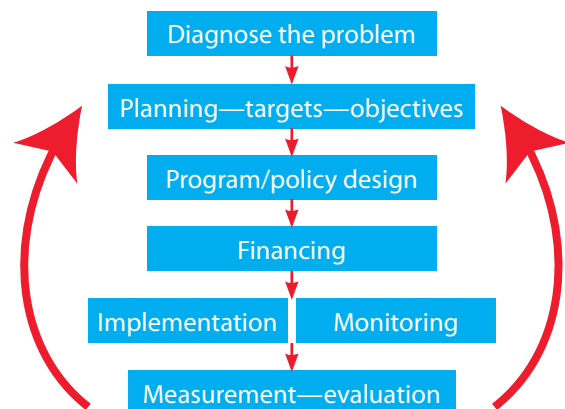
Figure 1 illustrates a results-based decision-making cycle, which ideally starts with a clear diagnosis of the problem that policy intends to address. Based on the diagnosis, the planning stage then defines policy targets and objectives. In the realm of social policy, these are normally welfare dimensions that can be modified through government actions.

Once the objectives are clear, specific programs or policies to accomplish them can be designed. When financing is available, implementation can begin and monitoring mechanisms put in place to ensure that the program is being executed according to plan. Once the policy action has taken place and the service or goods have been delivered, it is possible to measure whether the original objective has been accomplished and to evaluate whether goals have been effectively and efficiently achieved. Moreover, through evaluation it is also possible to compare the result with a counterfactual scenario of no policy, or of the use of alternative programs for achieving the same results, and therefore identify unexpected side effects or even negative impacts.

The process becomes a results-based cycle when measurement and evaluation results feed into the planning and policy design stages for fine-tuning, redirecting, changing the nature of the program or even eliminating it altogether if evidence shows that the program is not capable of delivering the expected results or produces unintended effects. The essence of the cycle, illustrated by the arrows in figure 1, is that the evidence generated is continuously and systematically used for policy improvement.

Interestingly, although the concept has been in vogue since the 1990s, when social expenditures started expanding in developing countries, particularly

Figure 1. Results-Based Policy Cycle

Diagnose the problem
Planning—targets—objectives
Program/policy design
Financing
Implementation | Monitoring
Measurement—evaluation

Source: Author's illustration.

in Latin America, a shared concern across the development community converged around the idea that too little evidence was available, and data that were available were seldom used for improvement. This concern was expressed eloquently by Savedoff, Levine and Birdsall (2006), who noted that foreign aid had increased by 70 percent between the years 2000 and 2005, with little evidence on its development impacts.

Since the mid-2000s, however, part of this picture started to change with a surge in the number of evaluations of different types performed on a variety of social policy interventions. Different methodological approaches have been followed including qualitative evaluations, impact evaluations, and summative evaluations, among others. As noted in the recent report by the World Bank's Independent Evaluation Group (2012), the Bank generated, on average, 16 impact evaluations of its programs per year between 1999 and 2004, while the average increased to 62 per year between 2005 and 2010. Other institutions have also increased the pace of their program evaluations. For example, the Inter-American Development Bank has performed 154 impact evaluations since 2008, compared to just a handful before that year (IDB 2013).

Specialized institutions were created to promote evaluation practice, including the Abdul Latif Jameel Poverty Action Lab (J-PAL) at the Massachusetts Institute of Technology in 2003 and the International Initiative for Impact Evaluation (3ie) in 2007, which have been able to attract considerable funding from public and private sources for generating hundreds of evaluations of social programs in developing countries over the past decade.[1]

The United Nations Development Programme (UNDP 2013) recently documented the methodological innovations proliferating in the social program evaluation field, which reflects the interest of practitioners, analysts, and academics in generating new evidence on what works for improving the standard of living for residents in developing regions.

However, much less progress seems to have been made on the other side of the equation regarding the *use* of evidence for improving policy. For example, the World Bank (2012) stresses that there is a growing challenge in ensuring systematic use of evidence even in the design of the institution's new projects and operations. Organizations such as 3ie have also started devoting resources to link evaluations with policy improvement plans to promote the use of the growing body of evidence.[2]

Different studies analyzing the "low-use" aspect in the social policy evaluation field converge on identifying pathways for extending the use of results, such as increasing evaluation quality in terms of its timeliness, credibility and clarity, and enhancing its relevance, improving its communication, and promoting an evaluation culture.[3]

These elements are certainly necessary to encourage policy makers to consider evidence during planning. However, there are at least two additional and less analyzed factors that are at play—incentives in the public sector and the way in which the evaluation function is organized.

Internalizing Evaluation into Policy

If evidence can improve the quality of policy, why hasn't use of evaluation results increased along with supply? One explanation is that policy decisions are made within a set of constraints and conditions that limit the incentives to generate and use information. For decision makers in the public sector, some of the relevant constraints and conditions include timing, the role of interest groups, politics, and inadequate technical capacities.

Political and evaluation cycles are often not synchronized because the window that governments have for delivering results is usually much shorter than the process for developing a solid impact assessment that requires design, baseline data generation, implementation, analysis, and identification of necessary changes. These stages can require years of investment that go beyond a political cycle and can discourage even strong advocates for evaluation.

Interest groups can also become an obstacle to implementation of rigorous evaluations, especially in the case of experimental designs. While the definition of control and treatment groups might be perfectly justifiable from a methodological point of view, explaining to program nonparticipants that they have been excluded from a perceived benefit because they were not "randomly selected," while others were, is not an easy task. This may create enough opposition to make conducting an impact evaluation infeasible.[4] In other cases, evidence of positive results may understandably generate pressure from control group members to be included, creating a threat of experiment contamination. Interestingly, pressures usually arise when maintaining the control group is more important, since the full, intended effect can only be measured when the intervention has time to yield its medium- or long-term impacts. The ability of evaluators to deal with interest group pressure usually determines the feasibility of impact evaluations.

Politics can present another obstacle. Generating information on the efficiency of policy action may be politically risky in some settings. While the government can capitalize on positive effects, unfavorable results may be much more difficult to handle and may require investing political capital for their management. The risk of obtaining negative or not-so-positive evaluation results might be a strong deterrent to promoting evaluation practice, and even more so if using the

evidence for fine-tuning implies reducing benefits to some groups, or possibly eliminating social assistance programs altogether. In these cases, the information might end up in a report with no impacts in the real world.

Negative results may also provide political opposition groups with ammunition that may backfire and become politically lethal, or may discourage external donors/ investors from allocating additional resources. The risk is usually higher in societies where strong transparency and accountability mechanisms are institutionalized and in environments of tight budget constraints, where many compete for resources (and where, paradoxically, information on more effective policies will be of greater value). Providing sound evidence on program impact under these circumstances may doom a particular action, rather than contribute to program improvement.

Policy analysts, advisers, and decision makers must have the technical capacity to understand evaluation results and use them to inform and improve future policy. The lower the professional capacities of these actors, the less likely it is that evidence will be translated into improvements in public policy. However, even in the context of low capacities, if evaluators are able to communicate their findings and conclusions clearly, the information is more likely to be used.

Evaluators can play an important role in facilitating the use of evidence. If they have sufficient independence and credibility, they can enhance accountability and transparency. When evaluation results are positive, this becomes a valuable asset that can be used politically to capitalize on successful government policies.

However, the objectives of the evaluator and the policy maker might not always converge. High-profile, academic evaluators may prioritize academic purity, professional prestige (academic success and/or high-profile publications), knowledge generation, and other factors that may be incompatible with timely, relevant, and communicable evaluations.

Program implementers are also key actors in the evaluation cycle, because they are responsible for operationalizing the changes based on evaluation results. Phasing in improvements that bring new standards, procedures, and methods into day-to-day operations can be a complex and laborious task, especially when the changes require modifications in individual behavior and customs. Even slight changes in procedures and norms can take several years for full implementation, and in practice, explicit or implicit opposition may inhibit evaluation use.

It is also common for civil servants to become "constituencies" of the program to which they have devoted years of effort. When program staff feel strong ownership, they can be the first to obstruct or openly oppose change, making it hard to effectively implement modifications based on evaluation evidence, even if at the higher decision-making level there is strong commitment to using evidence for policy improvement.

Finally, public opinion, political constituencies, and beneficiaries are also stakeholders in the process of producing and using evaluation results for improving social policy. Making information on policy impacts available enhances transparency and allows citizens to see how public resources are spent, or whether certain goals or benchmarks are met. Public dissemination also enables citizens to assess government performance and therefore acts as a tool to hold policy makers and program operators accountable. Transparency and accountability are highly valued in many political settings and constitute strong incentives for demanding evaluation production and use. Each actor involved in the social policy agenda plays an important role in influencing how evidence is generated and used.

Social Sector Arrangements for Evidence Generation and Use in Latin America

A country's governmental structure, particularly its social sector organization, is a critical factor in determining the extent to which evidence can be produced and used. The social sector in Latin America, for example, has a variety of arrangements: out of the 18 countries in the region for which information is available, 11 have created a social development ministry in charge of designing, coordinating, and implementing policies aimed at reducing poverty and inequality—9 of the ministries were established after 2002.

Figure 2 shows the number of functions of the results-based policy cycle (shown in figure 1) present in each country. According to the data, Chile, the Dominican Republic, Guatemala, Mexico, Nicaragua, Peru, and Uruguay have created units that are responsible for each of the elements in the results-based cycle and have a fully functional model. Ecuador has six of the functions—the missing element is the presence of a single institution focusing on planning and execution functions—while Argentina, Colombia, and Paraguay lack two of the elements. The remaining countries have four functions or less. Interestingly, the evaluation activity is absent in all the countries that have six functions or less.

As for the evaluation function specifically, Latin America employs three types of approaches. The first approach, the *centralized* model, has its monitoring and evaluation functions for the social sector assigned to an institution that does not belong directly to the sector. Sometimes these functions are concentrated in the country's office of the president (as in the Dominican Republic), while in other countries they are found in planning or finance ministries.

Figure 2. Elements of the Results-Based Policy Cycle Present in Each Latin American Country
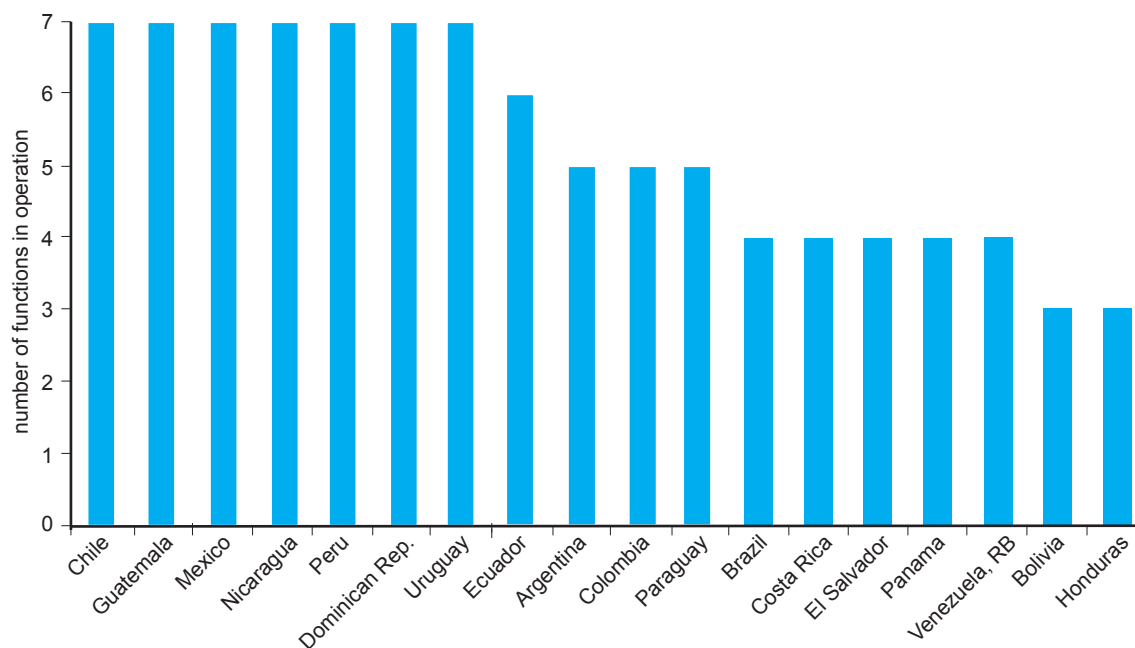
Source: Author's analysis.

These institutions or agencies have planning, monitoring, and evaluation responsibilities for all government sectors and focus mainly on budget disbursement and indicators included in national development plans. Their role focuses on generating information on the evolution of programs and policies for internal government use, rather than on preparing data for public dissemination.

There are pros and cons to the centralized model. An advantage of the centralized approach is that national government institutions have a general view for aligning government actions, even when they are executed by different ministries. Additionally, privileged access to detailed information on the operation of each program puts these institutions in an ideal position to identify areas for improvement. A downside, however, is that precisely due to their government actions, these institutions usually use similar evaluation instruments and approaches across all sectors, and are not able to tailor them to specific circumstances or needs. Therefore, the types of evidence produced under the centralized model are often too general to fine-tune program design and implementation, and the capacity of central government staff to follow through the process of improvement by other public institutions (for example, line ministries) is also limited. Given the tendency to highlight positive findings that imply government success, this type of information can be considered publicity rather than hard data that are useful for identifying areas for improvement.

Latin American countries that employ the centralized model include Argentina, Bolivia, Brazil, Chile, Colombia,

Costa Rica, the Dominican Republic, El Salvador, Honduras, and Uruguay.

The second approach, the *sectoral* model, assigns the evaluation function to an office that is part of a country's social development ministry. In contrast to the centralized model, these institutions are focused exclusively on social programs, which enables the use of evaluation tools specifically designed for the sector and provides a more in-depth analysis. In addition to generating information on disbursement and other operational indicators, these institutions are usually able to focus on measuring impacts and results.

A specific advantage of the sectoral model is that being a part of the same ministry responsible for program execution facilitates communication and feedback to program staff to influence policy planning and design. Since the evaluation function in these cases runs parallel to program implementation, it can provide an external view, while sharing the same objectives and goals of operational units, which facilitates the use of evidence for program improvement. A potential limitation of the sectoral model is that precisely because it shares the same line of authority with implementation units, its capacity for disseminating evidence is more limited, and the ministry can be biased toward highlighting positive results rather than pointing out areas for improvement.

Countries employing the sectoral model include Brazil, Chile, Guatemala, and Peru. Brazil and Chile are special cases because although they are included in this category due to the existence of an evaluation unit in their social

development ministries, they can also be categorized as the centralized model users because they have evaluation responsibilities in central government offices as well (Laguna 2011).

Finally, the third approach is the *independent* model, where the evaluation function is performed by an entity that is external to the social development ministry, but is dedicated exclusively to the evaluation of social programs. This reduces the conflict of interest that results from being part of the same institution and enables greater autonomy and transparency.

Under the independent model, evaluation activities are more likely to be unbiased and evaluators can more objectively direct their efforts toward identifying areas of improvement, rather than highlighting positive results. Independence also facilitates the dissemination of evaluation findings, which provides the public with impartial information on government performance.

An important feature of this model is that its independent nature helps build greater credibility for the evidence generated, as well as the resulting recommendations. A potential downside, however, is that being an external entity can also imply greater distance from the implementing agencies, and therefore, less influence on policy decisions. When evidence-based recommendations are not enforceable by the external agencies, their capacity for ensuring use of evaluation evidence is also limited.

Mexico is the only country in the region employing the independent model, which it does through its National Council for the Evaluation of Social Policy (CONEVAL). While CONEVAL is a government institution, it is led by a board of seven independent academic specialists with proven technical capacity that are not a part of the official government structure. In addition to carrying out independent evaluations, CONEVAL develops the norms and rules for ministries related to the social sector to use when conducting their own evaluations and coordinates the evaluation efforts of all actors involved in social policy execution.

Figure 3 highlights the institutional structure and advantages of each of the three models, as well as the challenges each of them faces in promoting the use of evaluation evidence in the social sector.

Good Practices in the Use of Evaluations: Evidence from Mexico, Peru, and South Africa

Several countries have made progress in closing the results-based management circle in the social policy realm.[5] This section discusses three country case studies where the different approaches to organizing the evaluation function (described previously) succeeded in influencing policy.

In South Africa, program managers were provided with a set of tools to improve their understanding and use of evidence. The Department of Evaluation in Peru used a different strategy based on building up capacities so

Figure 3. Three Models for Organizing an Evaluation Function in the Social Sector in Latin America

Model	Advantages	Challenges
1. *Centralized:* The evaluation function is allocated to central ministries (for example, finance, planning, or the presidency)—Argentina, Bolivia, Brazil,* Chile,* Colombia, Costa Rica, El Salvador, Honduras, the Dominican Republic, and Uruguay	— Access to detailed information — Control over budgets, resulting in coordination across government bodies and ability to influence program design and implementation	— Tendency to document positive government results — Limited capacity to assure changes in implementation — Use of nonspecific evaluation instruments (applicable to different sectors)
2. *Sectoral:* Evaluation function inside Social Development Ministry—Brazil,* Chile,* Guatemala, and Peru	— Generation of sector-specific information and analysis — Closer to executing agencies, which facilitates communication of feedback and results — Independence from policy implementation	— Can be perceived as punitive function, making internal coordination difficult — Potential tensions due to issues with academic rigor, practical use, and timeliness — Tension between transparency and political risks of negative results
3. *Independent:* Evaluation function in the social sector, but independent from the ministry structure—Mexico	— High credibility — Academic rigor — Transparency — Accountability	— Potential tension between academic rigor and relevance — Potential tension between transparency and political costs — Limited influence on program design and implementation due to external evaluation function

Source: Author's analysis.
* Can be classified in two categories.

that staff could work directly with implementers during analysis of evaluation recommendations and help develop an action plan for program improvement. The Mexico case study shows how a more independent evaluation body had considerable impact on policy priorities by communicating results to the country's president.

South Africa: Facilitating the use of evidence

The Department of Performance Monitoring and Evaluation (DPME) of the Presidency of South Africa was established in 2010 under the centralized model with the explicit objective to "*...strive for continuous improvement in service delivery through performance monitoring and evaluation.*"[6] To fulfill its mandate, DPME developed a set of norms, procedures, and guidelines for facilitating the adoption of evidence and its conversion into an input for program improvement in different line ministries, including the Department of Social Development. From the outset, DPME perceived its mission as contributing to development of better policies, where monitoring and evaluation are not seen as the ends, but rather as the means.

Three DPME instruments are of particular interest.[7] The first consists of guidelines for communicating evaluation results to program staff. These guidelines are to assist those in charge of generating evidence on the performance of government programs, and include two elements of support. The first element identifies the stages at which communication is important, such as when the evaluation is commissioned, implementation briefings, and dissemination of available results. The other support element details the communication mechanisms for sharing and disseminating evidence. These mechanisms include: (i) validation sessions with the main stakeholders to discuss results and obtain feedback from program staff and practitioners; (ii) discussions with program managers to get their feedback on the results; (iii) developing dissemination strategies targeting the relevant audiences—including organization of workshops, executive summaries and policy briefs, among other tools; and (iv) presenting results to the academic community to validate their methodological robustness.

The second instrument is a guide on how to develop a management response to an evaluation report, which is the mechanism for those in charge of policy implementation to establish their own position on the recommendations issued by the evaluator.[8] This is an important mechanism to promote the use of evidence, help those responsible for program operation to structure their views on the evaluation results, and identify areas for improvement in the evaluation process itself. Additionally, the guide offers templates and formats that facilitate the acceptance of the main findings as well as the identification of areas for improvement.

The third instrument assists program managers in developing an improvement plan to incorporate evaluation recommendations. This instrument is a key tool that helps program managers identify evidence to improve program performance. To assist in the development of the improvement plan, templates and formats are provided to record the improvement objective, establish expected outcomes, and define priorities, tasks, activities, necessary resources, and responsible actors. A timeline of implementation activities is also provided, including a systematic process for monitoring execution of the improvement plan. The sole objective of the improvement plan is to ensure the use of evaluation findings.

According to the latest DPME reports, these instruments have already been executed successfully in early childhood development interventions, schooling quality programs, and child support grants, all of which have been enhanced through the use of program evaluation evidence.

An important lesson from the South African experience is that (as clearly stated during DPME's creation) the generation of evidence alone was not enough for improving policy. If program managers are not supported and walked through the process of production, internalization, understanding, and use of the available evidence, the chances of closing the results-based circle can be undermined.

Peru: Managing evidence for policy improvement

In Peru, the *evidence management* function is assigned to the Evaluation Office of the Ministry of Development and Social Inclusion (MIDIS). The Evaluation Office was created in 2011 along with MIDIS, and was placed in the Vice Ministry of Social Policy and Evaluation, in line with the sectoral model described earlier.

Although the original objective of the Evaluation Office was restricted to evaluating social programs only, it was rapidly recognized that generating evidence would not automatically lead to policy improvement if program managers did not have the technical resources and time for absorbing and using the information. The Evidence Management Division was created to address these issues. The division seems to be the first of its kind in Latin America, and perhaps in the developing world. Its objective is to work with program managers to ensure that they receive information in a clear and user-friendly manner, and that they actually utilize it to develop an action plan for improvement.[9]

As a means for institutionalizing the process, the Evaluation Office modified the logical framework under which it had initially been operating to include the evidence management component explicitly. Figure 4 shows the current scheme. Following traditional setups, the logical framework includes the activities, results, and purposes of the evaluation function (first three elements in figure 4,

from the bottom up). Activities include dissemination, the generation of guidelines for quality control for evaluations performed by other agencies, evaluation services, analysis to identify recommendations, and training and advisory services. The expected results from these activities are inputs for social program implementers that will improve social policy performance.

Up to this stage, the scenario is similar to the traditional setup of the evaluation function in the public sector, where the purpose is to *contribute* to the improvement of policies and programs. By stating the purpose in this way, strictly speaking, the evaluation function would accomplish its objective by providing evidence to decision makers. Up to this point, the task of using the evidence for improvement is left in the hands of a third party—which is responsible for managing the program or policy action—so it is possible that even if the evidence is solid and relevant, in the end it might not be used at all.

The innovation of the MIDIS Evaluation Office is the inclusion of an additional dimension to the framework: the specification that the final objective is not generating evidence, but increasing "the impact, efficiency, quality, equity, and transparency of development and social inclusion policies and programs" (top element in figure 4). By introducing this statement explicitly, the objective of the evaluation function transcends the task of generating data, information and analysis, and goes

further by internalizing the goal of improving policy. Generating high-quality evaluations is only a first stage under this scenario.

To fulfill its ambitious and full mandate in a context of limited resources, the Evaluation Office established two lines of activity. One, which is similar to the recent efforts observed in South Africa, consists of developing guidelines, norms, formats, and procedures to facilitate the internalization of information by program managers, including instruments for communicating evidence in a structured and clear manner, as well as the development of an action plan by program implementers to address issues identified by evaluations.

The second line of activity, which is the main innovation introduced by the MIDIS, is that the Evaluation Office also provides support for internalizing the evidence, for developing and implementing the action plan, and for assessing the welfare implications of introducing the recommendations. The main result is that the measure of success of the evaluation function is modified. Normally evaluation bodies are assessed by the number of evaluations performed, reports generated, analyses offered, and so forth, which, in the MIDIS framework, would reach up to the third stage in figure 4. Under the evidence management approach, the extent to which the evidence impacts policies becomes the standard for assessing the contribution of the evaluation.

Figure 4. MIDIS Evaluation Logical Framework, Peru
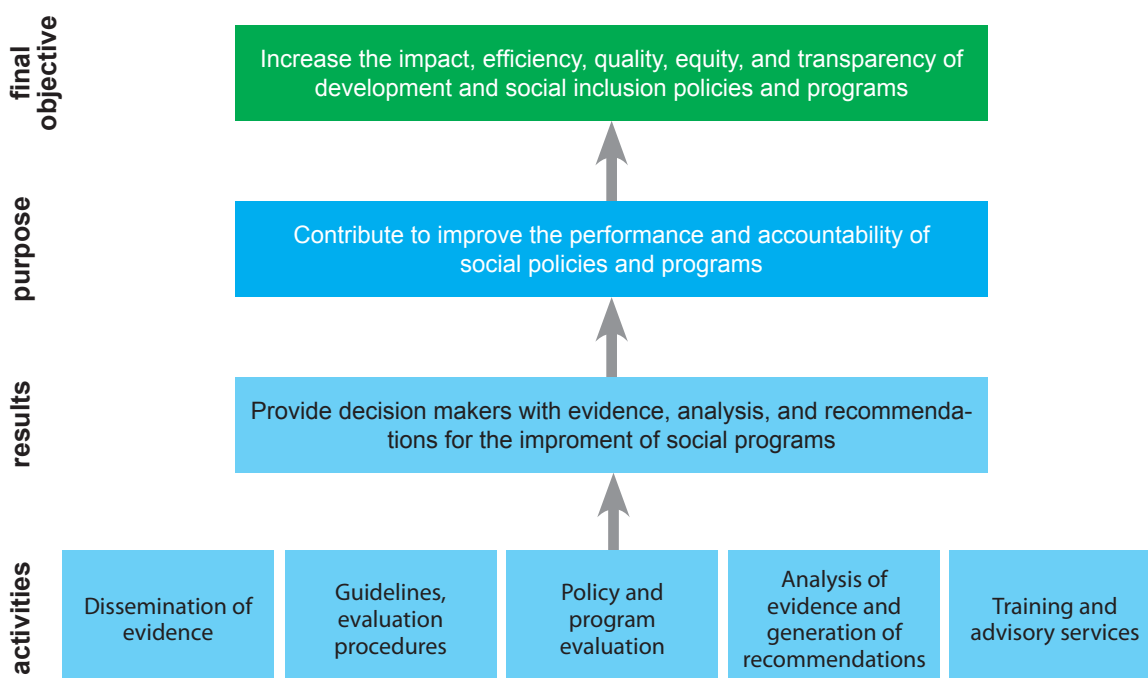
Source: Author's illustration.

Therefore, to assess the value added of evaluation under the framework in figure 4, it is necessary not only to ensure the use of evidence, but to also measure the consequences of doing so. MIDIS recently initiated a pilot exercise for performing such measurement, which is expected to be completed in 2014. This additional effort is a clear example of the benefits of implementing a results-based management cycle, which is currently under development in Peru.

Mexico: Communicating the results effectively

In contrast to the South African and Peruvian examples, CONEVAL works under a logical framework that only includes the first three lower levels in figure 4: activities, results, and purposes. Thus, the institution accomplishes its goals by making evidence available to decision makers, but is not responsible for guaranteeing its use—a contrast to the case of MIDIS, where this element is part of the evaluation function's mandate.

During the creation of CONEVAL, there was substantial discussion on the use of findings, and in particular on whether implementation of evaluation recommendations should be forced on program managers. It was decided not to include this responsibility based on the rationale that program managers are ultimately responsible for the success of the policy action and are more familiar with the context in which policy is implemented, so they should have the freedom to decide how to use the evaluation evidence.[10]

One of CONEVAL's achievements was effective communication of the impact evaluation findings of the Cement Floor Social Program in 2006, which resulted in program expansion and reduced Mexico's poverty by 10 percent. An impact evaluation indicated significant health, nutrition, and educational gains for household members when dirt floors were upgraded with a cement cover (Hernández 2014). The results were presented to President Calderón, who took office at the end of 2006. The government decided to increase the program's budget 100 times during the following six years based on program evidence. This increase improved housing conditions for 2.6 million families during this period.

In addition, access to cement floors was included as a component of the multidimensional poverty measure used to evaluate the country's performance, which incentivized local governments to support investment in similar approaches.

An interesting element of this example is that even though CONEVAL is only mandated to generate evidence, when it communicated results to the highest possible level, it was able to influence policy.

From Evidence Generation to Evidence Use

Over the last decade, the availability of evidence on what works in social policy has significantly expanded. However, for this evidence to play its expected role as a key input in results-based management systems, it has to be used systematically. At present, the challenge seems to be precisely that the growing body of available information is underused.

This is one of the challenges faced by Latin America and Caribbean countries. Although the region is often viewed as having relatively advanced monitoring and evaluation systems compared to other developing countries, in fact, less than half of the region's countries have adopted all of the elements of the results-based management cycle.

Among the underlying reasons for the slow uptake in the public sector, two are key. The first includes political, technical, bureaucratic, and operational factors that provide negative incentives for internalizing evidence systematically. The second reason relates to evaluation arrangements within governments that may lead to one of the following problems: the tendency to highlight positive outcomes in centralized and sectoral cases; tensions from academic rigor, practical use, and timeliness that can arise when evaluation and implementation units belong to the same sector; and difficulty in influencing program managers who do not belong to the same government agency in an independent model.

The examples of Mexico, Peru, and South Africa show, however, that it is possible to make progress in evaluation use for social policy improvement. In South Africa, which employs the centralized model, the strategy has been to develop a series of instruments that facilitate the understanding, analysis, and internalization of evidence by program managers to place the focus on improvement. In Peru, which uses a sectoral approach, the strategy has consisted of developing capacities in the Evaluation Department so that evaluation staff can support program implementers step by step in the identification and understanding of recommendations, and in the development and execution of improvement action plans. In Mexico, which is the only example of an independent model in the Latin America and Caribbean region, policy was influenced when the evaluation institution, CONEVAL, clearly communicated findings to the top decision-making level of government. These three experiences suggest that closing the results-based management circle is feasible, and that it is possible to do so in other developing countries and regions.

In the future, it will be important to regularly review the performance of the evaluation function in governments in terms of its success in improving efficiency and effectiveness in the public sector. This would allow the evaluation function to evolve into a results-based management tool itself.

About the Author

Miguel Székely is the *Director* of the *Center for Education and Social Studies (Centro de Estudios Educativos y Sociales)* in *Mexico City*. He was *Undersecretary for Planning and Evaluation* at the *Social Development Ministry* in Mexico between 2002 and 2006, and *Undersecretary* for *Middle Education* between 2006 and 2010. From 1996 to 2001, he was a *Research Economist* at the *Inter-American Development Bank*. He has a *PhD in economics* from the *University of Oxford* and has 76 academic publications to his credit on various development topics including evaluation, social policy, poverty, inequality, and education issues. He has lectured at the *Instituto Tecnológico Autónomo de México (ITAM)* and at the *University of Oxford*.

Notes

1. 3ie recently documented the surge in evaluations by integrating an inventory of 2,500 impact evaluations performed in 115 developing countries that met certain quality standards; most of these evaluations were performed after 2000. Another example was provided by McEwan (2013), who identified 76 impact evaluations using experimental methods focused on the determinants of learning. More than 80 percent of the available evaluations were performed during the 2000s.
2. For a closely related discussion on the conflict that exists among the different ways to use evaluations, see Briceno (2010).
3. See for instance, Läubli Loud and Mayne (2014), Quinn Patton (2008), and World Bank (2012).
4. Other methodological approaches, such as the use of quasi-control groups, can be used to avoid this potential problem, although they are not generally considered a "gold standard."
5. There are also interesting examples from institutions. One example recently promoted by 3ie is the use of "policy influence plans" as a requirement for generating evaluation designs for social programs.
6. See http://www.thepresidency.gov.za/pebble.asp?relid=1689.
7. The full National Evaluation Policy Framework, which explains all guidelines and procedures, can be found at http://www.thepresidency.gov.za/MediaLib/Downloads/Home/Ministries/National_Evaluation_Policy_Framework.pdf.
8. This is similar to performance audits that include management responses to the audit findings.
9. For more details, see http://www.midis.gob.pe/index.php/es/direc-general-de-seguimiento-y-evaluacion/direccion-de-evaluacion-de-politicas-sociales.
10. This contrasts with the Chilean design, where the Finance Ministry is active in selecting programs for evaluation from different ministries and follows up on the program improvement activities derived from the resulting evidence. Furthermore, the Finance Ministry supervises implementation of recommendations and publically reports the extent of their compliance.

References

Briceno, Bertha. 2010. "Defining the Type of M&E System: Clients, Intended Uses, and Actual Utilization." The Nuts & Bolts of M&E Systems, PREM Notes Special Series No. 2, World Bank, Washington, DC.

Hernández, Gonzalo. 2014. "Impact of (the Evaluation of) the Cement-Floor Program on Extreme Poverty: Mexico 2008–2012." CONEVAL, April.

IDB (Inter-American Development Bank). 2013. *IDB Development Effectiveness Report 2013*. Multilateral Investment Fund.

Laguna, Mauricio Dussauge. 2011. "Chile's Monitoring and Evaluation System, 1994–2010." The Nuts & Bolts of M&E Systems, PREM Notes Special Series No. 13, World Bank, Washington, DC.

Läubli Loud, M., and John Mayne. 2014. *Enhancing Evaluation Use: Insights from Internal Evaluation Units*. Los Angeles: Sage Publications.

McEwan, P. 2013. "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments." Wellesley College, Wellesley, MA.

Quinn Patton, M. 2008. *Utilization-Focused Evaluation*. Los Angeles: Sage Publications.

Savedoff, W., R. Levine, and N. Birdsall. 2006. "When Will We Ever Learn? Improving Lives through Impact Evaluation." The Evaluation Gap Working Group, Center for Global Development, Washington, DC.

UNDP (United Nations Development Programme). 2013. "Innovations in Monitoring and Evaluation." Discussion Paper. http://www.undp.org/content/undp/en/home/librarypage/capacity-building/discussion-paper--innovations-in-monitoring--evaluating-results/.

World Bank. 2012. "World Bank Group Impact Evaluations: Relevance and Effectiveness." Washington, DC.