

Behind the scenes: Managing and conducting large scale impact evaluations in Colombia

Bertha Briceño, Laura Cuesta and Orazio Attanasio
December 2011



About 3ie

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

3ie Working Paper series covers both conceptual issues related to impact evaluation and findings from specific studies or systematic reviews. The views in the paper are those of the authors, and cannot be taken to represent the views of 3ie, its members or any of its funders.

This Working Paper was written by Bertha Briceño, Water and Sanitation Program, World Bank, Laura Cuesta, University of Wisconsin-Madison and Orazio Attanasio, University College London

Acknowledgements

Luis Carlos Corral was advisor and provided guidance throughout the implementation of this study. The authors are also grateful to Howard White for his valuable comments. Generous support was provided by the International Initiative for Impact Evaluation (3ie). The views, findings, interpretations, and conclusions expressed in this article are entirely those of the authors and should not be attributed in any manner to the World Bank, its Board of Executive Directors, or the governments they represent; or the National Planning Department (Departamento Nacional de Planeación, DNP).

Photograph: Linda Epstein / istockphoto
3ie Working Paper Series Production Team: Radhika Menon and Mukul Soni

© 3ie, 2011

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

Contents

Abstract	3
1. Introduction	4
2. An ideal impact evaluation process	6
A. General considerations.....	6
B. Before starting the IE	7
C. During implementation of the IE.....	7
D. After implementing the IE	8
3. Facing the real world: learning from the experience of selected impact evaluations in Colombia	9
A. The impact evaluation of the Familias en Acción (FeA) conditional cash transfer program.....	10
B. The impact evaluation of the Jóvenes en Acción (JeA) youth labor training program	15
C. The impact evaluation of the Hogares Comunitarios de Bienestar (HCB) nursery program	18
D. The evaluation of the Sistema General de Participaciones (SGP) fiscal transfers' policy.....	22
4. Conclusions and key lessons	26
Annexe	32

BEHIND THE SCENES: MANAGING AND CONDUCTING LARGE SCALE IMPACT EVALUATIONS IN COLOMBIA

Bertha Briceño

Water and Sanitation Program, World Bank

Laura Cuesta

University of Wisconsin-Madison

Orazio Attanasio

University College London

Abstract

As more resources are being allocated to impact evaluation of development programs, the need to map out the utilization and influence of evaluations has been increasingly highlighted. This paper aims at filling this gap by describing and discussing experiences from four large impact evaluations in Colombia on a case study-basis. On the basis of (1) learning from our prior experience in both managing and conducting impact evaluations, (2) desk review of available documentation from the Monitoring & Evaluation system, and (3) structured interviews with government actors, evaluators and program managers, we benchmark each evaluation against eleven standards of quality. From this benchmarking exercise, we derive five key lessons for conducting high quality and influential impact evaluations: (1) investing in the preparation of good terms of reference and identification of evaluation questions; (2) choosing the best methodological approach to address the evaluation questions; (3) adopting mechanisms to ensure evaluation quality; (4) laying out the incentives for involved parties in order to foster evaluation buy-in; and (5) carrying out a plan for quality dissemination.

Key words: monitoring and evaluation, impact evaluation, best practices, case study, Colombia.

1. Introduction

The year 2010 marked the fifteenth anniversary of the Colombian monitoring and evaluation (M&E) system, formally the National System for Evaluation of Public Sector Performance (SINERGIA). During this period, substantial experience of commissioning evaluations of governmental programs has accumulated, and the focus of the system has evolved. It started with an ad hoc approach to evaluation, driven by opportunistic interests and support from multilateral institutions. Later, it began establishing a broader framework for M&E of public programs across the government with a particular focus on social interventions targeting low income populations (World Bank, 2009a).

During the inception stage, the system was conceptualized and focused on developing monitoring tools. Later on it incorporated evaluation activities, began to expand its scope to broader M&E approaches and played the role of an M&E system for the entire government. The National Planning Department (Departamento Nacional de Planeación DNP)¹ through one of its directorates has been the operational leader of the system. Although SINERGIA still lacks a formally defined architecture, clarity on funding sources, stakeholders and formal roles, it is considered a successful model in the region.

In the early 2000s, SINERGIA began supporting the implementation of impact evaluations (IEs) of three social programs that were launched during the recession of the late nineties. The first evaluation was commissioned to the consortium composed by Econometría (Colombian consulting firm), SEI (Colombian data collection firm), and the Institute for Fiscal Studies, London. This consortium won the first competitive bid for the evaluation of the Familias en Acción (FeA) conditional cash transfer (CCT) program. The bidding process started in the last few months of the economic downturn of 1999, when the Colombian government began to implement the CCT program with the hope of protecting the human capital investments of the poorest households. This first impact evaluation would later become the milestone for the implementation of IEs in Colombia, constituting a turning point in the development of the M&E system.

The IEs of the social network protection programs were followed by IEs of other social interventions such as Hogares Comunitarios de Bienestar (nursery program), and Vivienda de Interés Social (housing program), and CCTs for both displaced and indigenous populations, among many others. So far SINERGIA has coordinated 61 evaluations, spending US\$ 15 million in their implementation (DNP, 2010). According to statistics from DNP, 36% of these evaluations were conducted under an impact evaluation approach².

The particular origin of SINERGIA has determined different challenges for practitioners of IE in Colombia. One of the forces driving the increased use of impact evaluation has been the demand from the development banks (the World Bank and Inter-American Development Bank). The banks included the impact evaluations in their loans for social protection

¹ Government agency that is responsible for long term public policy design.

² Which represents 22 impact evaluations. It is not clear if these recent statistics follow the more rigorous definition of impact evaluations (experimental or quasi-experimental) originally adopted by DNP or a broader one

programs during the economic downturn in 2000. Before that, evaluations of public interventions were rarely conducted, and IE was not even considered feasible among the usual M&E approaches in Colombia.

Experiences from diverse countries practising impact evaluation suggest that some conditions facilitate institutionalization of IE and their utilization. Undoubtedly the Colombian M&E system has made important progress in establishing an evaluation culture, but still faces some obstacles to fully institutionalize its practice: the insufficient capacity of the government's agency responsible for managing evaluations, the weak buy-in among some key stakeholders, and the fact that the evaluation process is still not fully integrated into the national budget formulation and development planning process (World Bank, 2009a).

Although the Colombian M&E experience has been frequently highlighted as a relevant case for practitioners of IE, very little is known about how the practice of impact evaluation has evolved. Filling this gap is the motivation for this paper. We aim to present and summarize the experiences in the practice of impact evaluation from an implementation process point of view. To this end we rely on our experience in both managing and conducting IE in Colombia; using a case study methodology that involves a desk review of documentation from the M&E system and testimonies from sources with three different perspectives. The methodology included interviews with government actors, evaluators and program managers, with the purpose of deriving a complete view of the process.

As part of this study, we reviewed existing documentation of relevant cases, including methodological and final reports produced under the evaluation projects, and related documentation from other sources such as academia and the press. The desk review was complemented with semi-structured interviews of key stakeholders involved in the evaluations. The testimonies served in multiples ways: as a source of additional information, to contrast diverse perspectives in the process, and to identify lessons shared and common viewpoints regarding the implementation of the evaluations.

The rest of this paper is organized as follows: section two identifies key elements for carrying out an ideal impact evaluation. We highlight the importance of the independence of the evaluation process, quality controls, feasibility, and timeliness, among others. Section three summarizes the experience of implementing the evaluation of four major programs in Colombia: the Familias en Acción (FeA) CCT program, the Jóvenes en Acción (JeA) youth labor training program; the Hogares Comunitarios de Bienestar (HCB) nursery program, and the fiscal transfers policy Sistema General de Participaciones (SGP). This selection was based on budget allocation and policy relevance. The evaluation of all these except SGP was based on an exclusive IE approach. Given the magnitude and time of implementation of the decentralization policy, the evaluation of SGP involved a comprehensive framework, of which IE was one component. Section four draws lessons from managing these evaluations in Colombia.

2. An ideal impact evaluation process

Despite SINERGIA's limitations (insufficient institutional capacity, lack of own budget allocation, and limited autonomy, [World Bank, 2009a]), it can be argued that impact evaluation has gained credibility as a key tool for public management in Colombia. As of 2010, DNP reported around 23 ongoing impact evaluations amounting to US\$ 11 million (DNP, 2010). In general the impact evaluation practice in the context of SINERGIA has followed the rigorous definition of a comparison of outcomes of a program against an estimated counterfactual, through experimental and quasi-experimental designs. In the system, impact evaluation is formally defined as a causality analysis to determine effects of a program on beneficiaries (expected or unexpected, direct or indirect).

As the evaluation literature highlights, IE could be helpful in determining whether a pilot program should be replicated on a larger scale; IE allows the quantification of public interventions effects' and provides information to determine whether policy goals are achieved or not (World Bank, 2009a). However, IE is most useful only if it is conducted under certain conditions. Indeed, the same literature that highlights the benefits of IE points out the importance of following quality standards in order to carry out a successful IE.

In theory, several conditions to conduct an ideal impact evaluation can be identified. In this section we draw attention to some standards that guide useful, robust and unbiased IE. This selection draws extensively on the Quality Standards endorsed by the Development Assistance Committee (DAC), and the recommendations of the booklet "Institutionalizing impact evaluation within the framework of a monitoring and evaluation system" from the World Bank (2009a).

The overview includes general considerations for IE stakeholders within three broad phases of the IE implementation process: before the IE implementation (first phase), during the IE implementation (second phase) and after the IE implementation (third phase). This framework will facilitate understanding of challenges that emerge from managing IE in Colombia, which is the topic of chapter four.

A. General considerations

Throughout the entire implementation process IE stakeholders should pursue the achievement of the following standards (OECD, 2010; World Bank, 2009a):

- 1. Independence of evaluation process:** To ensure the credibility of IEs' results, evaluators should be independent from program/policy management. Likewise, evaluators should be able to make free assessments without undue interference from managers or other stakeholders.
- 2. Contribution to capacity building and promotion of evaluation culture:** IE should foster opportunities for evaluation capacity building inside the programs, knowledge transfer and evaluation promotion directed to maintain buy-in from stakeholders.

3. **Quality control:** This refers to the supervision of evaluators' work and guaranteeing rigor in the technical approach used. OECD (2010) suggests implementation of mechanisms such as peer-reviewing, advisory panels, or reference groups. An important aspect of quality control is compliance with ethical safeguards; an ideal IE should guarantee protection of human participants and follow the principles of respect, beneficence and justice. Particular attention should be given to acquiring informed consent, acquiring the necessary permissions and providing compensation when appropriate.
4. **Buy-in from stakeholders:** An ideal evaluation enjoys support from major stakeholders, especially from program management. The evaluation scope takes into account the demands and concerns from program managers, and there is expectation that the findings will be relevant and useful for improving performance and lesson-learning. Evaluators' expertise and understanding of the context are requisites for this. However, buy-in should not conflict with evaluator's independence. A rule of thumb is that discussions are centered on clarification of facts, additional information, methodological caveats or evaluation questions. Environments where program managers have low incentives for lesson-learning and where reporting failure is associated with poor performance are more conducive to undue interference.

B. Before starting the IE

Before the implementation process, IE stakeholders should pursue the achievement of the following standards (OECD, 2010; World Bank, 2009a):

1. **Clarity of the evaluation's object and scope:** Assuming that the decision to carry out an evaluation has already been made, the object of the evaluation should be unmistakably defined (type of intervention, program's objectives, components, population targeted, etc). The program's objectives are particularly important because they determine the outcomes on which the impact evaluation is going to focus and the overall IE scope. Likewise, the evaluation's terms of reference (TOR) should clearly define the scope of the evaluation, including its objectives and the evaluation questions, as well as the resources available, the time framework for IE implementation, the methodologies to be used, and a clear theory-based approach.³
2. **Feasibility of the evaluation:** Once the evaluation object has been defined, evaluation and program managers should decide the technical approach that best suits the needs of the evaluation's questions and objectives, assess if an impact evaluation is viable from the technical viewpoint, and if it is the best instrument to meet the concerns of policymakers. Viability of the IE will crucially depend on the ability to identify a credible counterfactual scenario, usually but not exclusively by the identification of treatment and comparison groups (under experimental or quasi-experimental designs), but also via other rigorous methodologies (simulations).

C. During implementation of the IE

During the implementation process, IE stakeholders should pursue the achievement of the following standards (OECD, 2010; World Bank, 2009a):

³ For a broader discussion see White (2009) and World Bank (2006).

- 1. Active interaction among stakeholders:** Ideally, evaluators should have a fluent contact with program managers and stakeholders, via clear channels of communication and contact points. Stages at which feedback is provided to high-level policymaking should be clearly defined and established. Interaction is important for keeping track of the evaluation development, for ensuring availability of necessary information, but also for understanding the key assumptions, the theory of change of the program, the findings and for facilitating the dissemination process as well.
- 2. Answers to evaluation questions:** the evaluation should satisfy expectations regarding the questions that motivated it. This entails accomplishing the evaluation's goals, and clearly documenting findings in the final reports. In addition, a crucial aspect for an ideal impact evaluation is that the evaluation should be guided by an explicit theory of change that maps the causal chain from inputs to impacts, sheds light on the reasons for absence or presence of certain impacts, heterogeneity of impacts, heterogeneity of treatments, and ultimately, helps assess the assumptions behind the interventions.
- 3. Strong relationship between findings and recommendations:** Conclusions and recommendations from IE should be derived from the study's findings. Clear evidence should support conclusions and recommendations. In addition, IE should explicitly state key issues and research questions worth further study. An ideal IE should constitute the basis for ex-post rigorous economic evaluation analyses (cost benefit or cost effectiveness), if collection of cost data is envisioned. Solely assessing impacts can be of limited policy relevance if there is no information generated on the cost of achieving those impacts (World Bank, 2006).

D. After implementing the IE

Once the IE is finalized, IE stakeholders should pursue the achievement of the following standards (World Bank, 2006):

- 1. Quality dissemination:** Evaluation findings are presented to all program stakeholders, especially high level officials with decision-making powers, and in general to all interested parties. This should be done in understandable language and a clear and concise manner, with frank reporting of both successes and failures, avoiding positive results bias. In addition, the IE should be envisioned as a public good; products derived from the IE such as data sets, programming algorithms, etc. are made available for further use, research and external validation (replication).
- 2. Timeliness and utilization of evaluation's findings:** Evaluation findings should be reported in time and should be used for accountability purposes. If the motivating questions were clear, and the evaluation's reports incorporated answers to these questions, the program's managers can use the findings to discuss recommendations and to support changes in the program, if necessary. The ideal evaluation process does not end with a published document; an ideal evaluation should document the recommendations' discussions, the response from program managers and the consequent decisions or improvement aspects resulting from the evaluation.

3. Facing the real world: learning from the experience of selected impact evaluations in Colombia

In the late nineties, Colombia experienced the worst recession since World War II. The crisis badly affected poor households in particular, with their income falling by about 35% between 1998 and 2000. Policymakers perceived traditional interventions to be both inadequate and insufficient for protecting vulnerable families from the effects of unemployment and lack of income (DNP, 2004). Some studies highlighted that resources for those interventions depended on the country's macroeconomic performance, and during the crisis there was limited capacity to respond by expanding coverage, or even maintaining the number of beneficiaries already included (Núñez & Cuesta, 2006; DNP, 2004).

In this context, the government launched three new social protection programs under the so called Plan Colombia with the explicit objective of offsetting the effects of the economic crisis. As part of this policy, in 2000, the government introduced the Red de Apoyo Social (RAS), a safety net strategy to protect the low income populations from the effects of the crisis. The RAS was financed with loans from the Inter-American Development Bank and the World Bank and its rationale was to implement new interventions that, in contrast to traditional ones, would offer the flexibility, efficacy and timeliness that the government was looking to counteract the social impacts of the crisis. The RAS looked at recent successful experiences in Latin America to inform its design. It is interesting to observe that some of these programs, as we will mention, had already been rigorously evaluated.

At its inception, the RAS included three major programs: (i) Familias en Acción (FeA), a conditional cash transfer program for poor families that would give cash assistance to those who kept their children in school and provided them with basic preventive health care; (ii) Empleo en Acción (EA), a community works program designed to provide temporary employment to poor, unemployed and unskilled workers; and (iii) Jóvenes en Acción (JeA), a training/apprenticeship program for young adults provided by private sector training institutions that were selected through a competitive bidding process.

Besides complementing the objectives and social policies established in the National Development Plan 1998-2002, RAS' programs involved important changes to the Colombian framework of social policy. First, they introduced the conditionality of subsidies delivery (cash or in-kind) for the accomplishment of some socially desirable goals: school attendance and medical check-ups (for Familias en Acción), job training (for Jóvenes en Acción) and work (for Empleo en Acción). Second, the interventions were established outside the traditional institutions responsible for running social programs. Third, the interventions were completely funded (for the initial period) by multilateral agencies' loans. This last feature also had implications for the evaluation of the new programs: the development banks included as one of the conditions of the loans that a fraction of the same would be used to evaluate the impact of each of the programs (DNP, 2004).

While the three evaluations were tendered internationally to guarantee independence and transparency, the start of the evaluation work also had important institutional repercussions. In 2000, the Colombian government lacked experience in managing impact evaluations. To build capacity, the government adopted different steps such as creating a special group within the DNP responsible for managing the IEs, hiring some evaluation experts to lead the process, and getting technical support from multilateral banks. In 2002, the special group was integrated into the Directorate for Evaluation of Public Policies and Management for Results (DEPP), which had been created in the late 90s. The RAS' IEs,

especially the FeA impact evaluation, would later underpin SINERGIA's evaluation pillar positioning IE as a credible and useful M&E approach.

It has been 10 years since the Colombian M&E system began to accumulate knowledge in managing impact evaluations. The experience has been mixed. Some of the evaluations conducted have been highlighted as successful examples for the region, in terms of the quality of the evaluation itself but also in terms of the influence on the program's design. Others, although of high technical quality and rigor, were archived without being referred to in any meaningful discussion or influencing the relevant policies. And of course there are also cases of evaluations that were not characterized by high scientific standards but nevertheless prompted important policy changes.

This section summarizes experiences and derives lessons from four selected evaluation processes carried out in recent years under the framework of the Colombian governmental M&E system, SINERGIA. Its purpose is to extract lessons from typical situations that occur in the evaluation practice, making the Colombian experience a relevant case for practitioners of IE. Table 1 in the Annexe presents a description and main features of each intervention.

A. The impact evaluation of the Familias en Acción (FeA) conditional cash transfer program

Familias en Acción is a conditional cash transfer program inspired by the PROGRESA/Oportunidades experience. It was the largest of the three initial programs included in the RAS. It was piloted in 2001 and its launch in a set of small towns in Colombia was planned for early 2002. After a long international tender in the second half of 2001, the Familias en Acción IE was awarded to a consortium composed of The Institute for Fiscal Studies of London, Econometría (Colombian consulting firm) and SEI (Colombian data collection firm), called the IFS-Econometría-SEI consortium.

The conclusion of the tendering process coincided with the end of the presidential period 1998-2002, and the incumbent in the President's office was not interested in the evaluation and did not even understand its purpose (Attanasio, 2009a). In the end, it took several months before the evaluation could actually start.

According to stakeholders interviewed for the study, the negotiations on the terms of reference (TOR) and on the contract between the President's office and the Evaluation team at DNP were so complicated that at some point the IE managers thought the process was going to fail. On one hand, there had not been a precedent for an evaluation of this type and scope in the country, and above all, some government officials thought the cost was simply scandalous. On the other hand, the DNP, through the special group commissioned to managing IEs, strongly advocated for the evaluation of RAS programs, pointing out the successful experience of Progresas-Oportunidades, the Mexican CCT that inspired FeA. The negotiation started in October 2000 and the contract was finally signed on the last business day of December 2001.

Establishing the TORs and signing the contract with the evaluators was only the beginning of a number of difficulties. In January 2002, the evaluation team started work on the evaluation design in collaboration with the evaluation team at DNP and with the direction of the program. The proposal designed by the evaluators contained two alternative plans. The first, and preferred one, wanted to use the expansion phase of the program for evaluation and envisaged the random assignment of the municipalities targeted by the program to

different starting points. The second plan, anticipating the political objections to a slowdown of the expansion phase in some municipalities, envisaged the use of a number of municipalities excluded from the program for a variety of reasons as a comparison group. As this plan recognized the possibility that excluded municipalities would be different from the included ones, it relied on the possibility of having a pure baseline survey to control for such differences.

Since randomized selection of beneficiaries was vetoed from the beginning, the consortium proposed the randomization of the expansion, as it was done in the Progresía-Oportunidades IE in January 2002. Unfortunately, the president's office also vetoed this type of selection arguing that elections were coming up in March 2002 (Attanasio, 2009a). It was clear that they would have never accepted such a design and did not understand its purpose. At that point the evaluation consortium with the support of the DNP IE team started to work on the second plan and on the identification of a group of municipalities that could be used as a comparison group.

FeA was targeted at municipalities that satisfied the following criteria: they had less than 100,000 inhabitants and were not departmental capitals, had enough infrastructure in health and education and had at least one bank branch. The evaluation group first stratified the universe of targeted municipalities according to two criteria: geography (five regions were identified) and an index of infrastructure. Within each of the resulting 25 cells, about two municipalities were drawn. The control municipalities were drawn from the same cells and, in practice, often resulted in municipalities that had been excluded because of the lack of a bank or, in some cases, because of population size limits. In early 2002, less than three months after being awarded the contract, the evaluation team had a sample of 57 treatment and 65 control municipalities and started to prepare the collection of the baseline data.

Unexpectedly, new obstacles to carrying out the new IE strategy emerged. This time the obstacle was the government interest in FeA expansion, which risked jeopardizing the possibility of collecting a pure baseline.

The consortium planned to start the collection of baseline data in June 2002, and finish it by October 2002. Surprisingly, in the early months of 2002, the President's office told the IE managers that they would start payments in all treatment municipalities immediately, which consequently prevented the collection of a pure baseline (Attanasio, 2009a). After several meetings, in which the officials of the evaluation departments at DNP, the IADB and the World Bank played a key role, the government accepted to keep about half the 'treatment' towns out of the program until the baseline was completed. As a consequence, the treatment sample was effectively divided into two groups, treatment without payment, comprising towns where enrolled families had not received the first payment; and treatment with payment, comprising towns where beneficiaries had received the first payment at the baseline data collection. Interestingly, which towns entered early and which towns entered late was, to an extent, random and mainly determined by the readiness of the administrative procedures to start the program.

This situation afforded a new opportunity that the evaluators seized: the fact that some towns would be in the program at the time of the first survey and others not, and that the early towns were selected more or less randomly (although not in a rigorous fashion), could be used to have some early impacts estimates immediately after the first data collection (although on a sample of a limited size). Given this opportunity, the evaluators added retrospective modules to the basic survey, which proved crucial in all phases of the evaluation.

Just when IE managers and evaluators had thought that the obstacles for IE implementation had been overcome, a change of President and the transition toward a new style of government, with new officials and new social policies, took place. From August 2002 to January 2003, IE managers and evaluators tried to convince the new President to maintain the program and support the evaluation. Although the early dissemination of results is one of the prominent achievements of the FeA impact evaluation, at that point in time, the consortium did not have evidence on the intervention's benefits which made negotiations more difficult.⁴

The new government had concerns regarding the cost and usefulness of the IE. In particular, the President's office was wondering if it was worth investing US\$ 2 million in the study, instead of using these resources to expand the program coverage. Again, negotiation with the President's office became so complicated that IE managers thought they would have to cancel the contract and the evaluation. Fortunately, well known and experienced economists mediated to support the study and helped change the President's stance. In this context, having the possibility to show some preliminary evidence on early impacts that exploited the fact that the program had already started at baseline in a subset of municipalities was particularly useful.

The results, which were quite convincing for some outcomes where the evaluators could exploit the presence of retrospective information (such as school enrolment), were included in the baseline report that was delivered to DNP in December 2002 (just a year after the signing of the evaluation contract!). The early impacts (that were later confirmed by the first follow up collected in the summer of 2003) were positive and were probably influential in achieving the continuity of the evaluation. Finally, in January 2003, the President gave his approval for continuing the IE of RAS' programs.⁵

As mentioned, the baseline was collected in 2002 and the first follow up completed in 2003. The evaluation surveys were large and complex, including a large number of variables and several questionnaires administered to several key agents. The household questionnaire was administered to about 11,500 households in 2002. A remarkable 94% of these households were re-contacted in 2003. The household questionnaire was complemented by school and health centres' questionnaires, questionnaires of local authorities and so on.

The total cost of the FeA to the Colombian government was about US\$2 million, including three waves of the questionnaires. This figure compares very favourably with other evaluations, such as the PROGRESA one, where a similar cost did not include any data collection and where the cost was borne by the Mexican government. In the case of FeA, the evaluation consortium secured additional funds to improve the quality of the data collection on three occasions. In 2003, an additional amount of US\$200,000, from the IADB, was used to limit attrition and ensure that most households would be re-contacted. In 2005-2006 an additional £25,000 (donated by the department of Epidemiology at UCL) was used to supplement the data collection with measures of haemoglobin among children and an additional £75,000 from the ESRC in the UK was used to construct measures of social capital using experimental games.

As mentioned before, one of the key features of the FeA IE was the prompt dissemination of its results. Early findings, based on the analysis of baseline and follow up data, came out in July 2004 and were quickly presented to a large audience, comprising government authorities, program managers, academia and the press. IE managers and evaluators who

⁴ By the time this happened, the consortium was collecting the data for the base line.

⁵ Five months after the beginning of the new government.

were interviewed for this report highlighted that the timeliness in findings' dissemination had impacts on both IE management and program implementation. Regarding the impact evaluation, stakeholders mentioned that it was instrumental in conveying support for the later phases of the evaluation.⁶ As for program implementation, evaluators pointed out that prompt dissemination changed the government agenda regarding the term of the intervention (Attanasio, 2009a).

Following the first follow up, a second follow up was conducted between 2005 and 2006. This allowed for the measurement impacts after three years of the inception of the program. The results of this additional evaluation exercise were presented in Bogota in October 2006.

The main results of the FeA impact evaluation showed significant improvements in children's school attendance and total consumption. The enrolment of children between 12 and 17 years increased by five percentage points in urban areas and seven percentage points in rural areas. Likewise, total consumption rose by nine percentage points in urban areas and eleven percentage points in rural areas. Food intake increased by 11 percentage points in both urban and rural areas. As for the impacts on health practices, the consortium found that medical check-ups among children from 48 months to 84 months increased by 38 percentage points and vaccination increased by 4 percentage points. In rural areas, the prevalence of diarrhoea among children under 36 months decreased by nine percentage points.

The evaluation reports (and in particular that of 2006) stressed that while FeA could be declared a success in some aspects, there were some problems in other areas. In particular, the impacts were much stronger in rural isolated areas and much weaker in 'cabeceras municipales', that is in the urban centres. Almost no impacts on nutrition and health outcomes were observed and the education results were much smaller. These limitations raised concerns for the expansion of the program to urban areas.

FeA soon began one of the flagships of the Colombian government social policy which was expanded considerably; in particular it was targeted towards the displaced population and at poor urban areas. The IE of the rural version of FeA was therefore followed by the IE of the other versions of the program: FeA directed toward displaced population⁷ and FeA in urban areas (including the IE of the pilots of Pasto and Medellín). Although one might think that at this point the obstacles to the evaluation would have been overcome, evaluators of both the pilots of Pasto and Medellín and more generally of the urban expansion also faced barriers for conducting an unbiased and rigorous IE.

Originally, an experimental design was considered for the IE of the urban pilot. The evaluation team of DNP identified two neighbourhoods in Soacha, a large satellite city of Bogota and was planning to use one as a 'treatment' and one as a comparison or 'control'. Moreover, in the treatment neighbourhood, different households (randomly selected) would receive slightly different versions of the program. However, once the baseline was completed, program managers did not comply with the commitment of maintaining the comparison group. Based on stakeholders' testimonies, there was pressure from the President's office to expand the program and achieve complete coverage in the municipality of Soacha. Apparently, this was seen as an opportunity to offset the reluctance to FeA implementation within the capital of Colombia. FeA has been seen as a flagship program

⁶ The *Familias en Acción* program directed toward displaced, indigenous population and urban areas.

⁷ Commissioned to the *Centro Nacional de Consultoría* (Colombian consulting firm) after an international tendered process.

from Uribe's administration, politically distant from the authorities of Bogotá. Finally, the program was expanded and the randomization managed at the baseline was lost.

Similarly, IE managers were forced to change the original design for the IE of the urban expansion. Initially the evaluation was based on a design that randomly assigned similar municipalities into two groups. This proposal aimed at evaluating different modalities of scholarships, in order to establish whether incremental increases in the amount of the subsidy or the provision of prizes for promotion and graduation were preferable to the traditional payment schedule in terms of educational outcomes. Unfortunately, after the baseline was completed, Colombia experienced a proliferation of Ponzi schemes that created social unrest. In an attempt to offset this situation the President ordered the acceleration of the introduction of the program in the cities selected for the IE of the urban expansion. Therefore, program implementation affected the IE implementation and the initial design had to be further adjusted to the new implementation circumstances. Once again, the collection of a valuable baseline was 'wasted' because of the political 'necessity' to expand a popular program quickly. As of today, no strong and systematic evidence exists on the impact of FeA in urban areas.⁸

To sum up, the experience of the evaluation of FeA, which was the first large scale evaluation of a social intervention implemented in Colombia, was on the whole, positive. Although it faced many obstacles, this IE had a tremendous demonstrative effect on establishing the relevance of evaluation of public policies in Colombia. Program managers obtained real gains from having engaged in the evaluation. Indeed, the positive results from IE convinced the President's office of the importance of giving continuity to the program instead of taking advantage of the potential political gains from launching a different intervention. Furthermore, based on the proven results and success of this program, many countries in the region and beyond adopted similar mechanisms for improving human capital development of the poorest segments of their population. Lastly, but not less importantly, the IE provoked an interesting debate among program managers, politicians, academia and the press regarding the effectiveness of CCTs in reducing extreme poverty in Colombia (Box 1).

On the negative side, one has to register the persistent resistance to implementing randomized evaluations, which is still relevant. Moreover, one cannot help notice that the executive appeared to react positively to the evaluation findings when these were aligned with its policy choices; but one wonders what would have been the reaction if the evaluation had found negative results. The executive certainly was not very reactive to the words of caution expressed by the evaluators about the urban expansion and was not particularly supportive of a rigorous evaluation in urban areas.

Evaluations can be particularly useful in the design stages of a new program. In this respect the urban expansion of FeA had great promise and it was interesting that, at some point, the program administration was willing to experiment with different modalities of the program in urban areas and to evaluate their impact. Unfortunately that commitment did not survive the evolution of political needs.

⁸ An exception are two studies of DNP on the pilots of Medellín and Soacha that use the available data in the best possible way and find some positive impacts of FeA in these contexts.

Box 1. Tracing the influence of evaluations: *Familias en Acción* in media and political contexts.

Familias en Acción cannot be considered the main social policy instrument of the government in Colombia. Its implementation has however motivated more op-eds and policy debates than any other social intervention in Colombia. Not surprisingly, the IE was not exempt from media and political attention. Indeed, the FeA impact evaluation became instrumental for the government's response to critiques from academia and political detractors.

Since its inception, *Familias en Acción* created many expectations among policymakers and some skepticism within academia. Based on the success of *Progresar-Oportunidades* in Mexico, government officials thought that FeA would be an appropriate policy for confronting some of the effects of the economic downturn, and later on, would become a central policy against poverty. On the other side, academia was concerned about the incentives that the program would create among low income populations.

Early critiques from academia were focused on the statement that *Familias en Acción* would exacerbate dependence on public assistance by increasing program participation, reducing labor supply, and therefore, discouraging beneficiaries from getting out of poverty (e.g. *El Espectador*, August 26th, 2006 and October 14th, 2006). Nonetheless, both the IE conducted by the consortium *IFS-Econometria-SEI* and the IE commissioned to the *Centro Nacional de Consultoría* showed that the program does not have any impact on labor outcomes, apart from discouraging child labor supply.

Later on, critiques were focused on the fact that *Familias en Acción* subsidies' are given per child. Hence, if the family had more children, they would receive more resources. Although there was no evidence of the program's contribution to increase in youth fertility rates, academics claimed in different op-eds that *Familias en Acción* had encouraged poor teenagers to get pregnant and become program's beneficiaries (e.g. *El Espectador*, August 18th, 2007; *El Tiempo*, May 4th, 2007). The argument was rigorously rejected by the IE conducted by the consortium *IFS-Econometria-SEI*. The latest critiques have been with regard to the urban expansion of the program against the advice of the IE (*El Espectador*, April 5th, 2008).

The IE has also been instrumental in protecting the government from critiques of different Colombian political actors. Moreover, stakeholders' testimonies indicate that it was crucial to demonstrate due diligence of government officials regarding the needs of displaced population before the judicial authorities. One of the best known cases of this fact is the utilization of the findings from the IE of the version directed to displaced populations as evidence of government's efforts in improving the critical condition of these families. Evaluators presented the findings in front of Constitutional court magistrates and managed to convince them of the program's beneficial impact on displaced households.

B. The impact evaluation of the *Jóvenes en Acción (JeA)* youth labor training program

After a new tendering process the impact evaluation of *Jóvenes en Acción (JeA)* was awarded to the same consortium that won the contract for conducting the *Familias en Acción* impact evaluation. The program was implemented between 2001 and 2005, and the

IE was conducted from 2004 to 2006, based on the fourth (and last) cohort trained under this intervention.

In 2005, the year the baseline data were collected, there were 118 training institutions and 1000 companies participating in Jóvenes en Acción. The program was implemented in the seven largest city of Colombia, at a cost of US\$700 per participant.⁹ An interesting feature of the program was the fact that the training institutions chose and designed the training courses they offered. They had to also guarantee an important on-the-job component. This implied that participating firms had to be willing to take the trainees on a subsidized contract for at least a few months. Moreover, the training institutions were given clear incentives to complete the training of the individual beneficiaries. Arguably, this design made the training program different from other existing programs especially in the decentralization of the choice of courses that linked it to the demand on the labor market (through the on-the-job component).

The Jóvenes en Acción impact evaluation had two features that were particularly unique within the Colombian M&E system: (i) the evaluation design was based on a randomized selection process of beneficiaries and (ii) the program being evaluated was a large one which was already operating at scale. The fact that it was feasible to randomize at the individual level (as discussed later) did not preclude the conduct of the evaluation even after the program had started (Consortium Econometria-IFS-SEI, 2007). The evaluation of Jóvenes en Acción is so far the first and only time that IE has been conducted as a pure experiment in Colombia¹⁰.

The design of the evaluation was particularly interesting because special care was taken to avoid biases arising from beneficiary selection by training institutions, particularly since the program at scale, effectively involved some selection of potential beneficiaries by the training institution. These features make this experience a particularly relevant case for IE practitioners, as well as a model of an ideal IE in methodological terms.

Jovenes en Acción was the last intervention of the RAS' programs. From 2002 to 2003, it was managed with support from the development banks by Acción Social (AS), the same agency that was implementing Familias en Acción. Just like Familias en Acción was inspired by Progresas-Oportunidades, the design of Jovenes en Acción followed the successful structure of a program called Chile Joven, supported by multilateral agencies in the region.

Unlike Familias en Acción, Jovenes en Acción was competing with traditional programs within the Colombian government architecture. The SENA, a governmental entity responsible for the execution of training and labor searching policies, viewed JeA as a threat to its current interventions, and pushed for the program to be relocated. While SENA was the public provider of training programs, JeA worked through private accredited training providers known as ECAPs (in Spanish, entidades de capacitación acreditadas). After some months of discussions between Acción Social, the Ministry of Social Protection, the SENA and the office of the new president, Jóvenes en Acción was inserted into the traditional institution responsible for training and labor programs.

The first and the second cohort were trained under the institutional framework of Acción Social, and the third and the fourth were instructed under the umbrella of the SENA. The IE was conducted among the fourth cohort. The evaluation design worked as follows. The

⁹ A better description of the program is in the Annex (table A-1).

¹⁰ At the time of writing the evaluation of the Red JUNTOS that is being conducted is based on a randomized Assignment design.

courses offered by the training institutions were designed to have a cap of 30 individuals. The trainees were chosen among the applicants by the training institutions, which had a strong incentive to select the better applicants, as they were paid on the basis of completed and successful training. After many negotiations the training institutions involved in the evaluation agreed to admit 45 applicants per course. The evaluation team selected the 30 trainees among these 45, keeping the remaining 15 as controls. Notice that this mechanism preserves the spirit of the system by which the training institutions select the best applicants, while at the same time retaining the comparability between treatment and control samples since the final allocation to the training courses is random.

The impact evaluation showed very positive effects on the labor market outcomes of its beneficiaries. Monthly salaries increased by US\$ 11, which is 12% higher than the wage earned by the control group. Likewise, evaluators found that impacts on women were extraordinarily positive. Participation in JeA increased women's wage earnings by 18% and formal earnings by 31% (Attanasio et al, 2009b). The qualitative analysis that accompanied this IE showed that women were more motivated during classroom and internship phases of the program, which may be one of the reasons that explain the remarkable impacts among women. The data from the evaluation was used in an academic paper which is forthcoming in the American Economic Journal (see Attanasio, Kugler and Meghir, 2011).

It is paradoxical that although Jovenes en Acción was a rigorous and unbiased impact evaluation, and the results showed very positive effects from the intervention, neither the government officials nor the program's managers exploited the potential of the IE results. It can be said that in this case the IE results did not affect the conduct of policy. When the IE results came out, the original program had already been turned into a new intervention, which only shared the name with the original program. New age limits and new rules regarding the training (provider, type of courses, etc.) were introduced. In particular, the most interesting feature of the program, the decentralization of course design and content to training institutions which had a direct link to the labor market and incentives to match the supply of specific skills to its demand was eliminated as the program was absorbed within the centralized structure of SENA.

There are many reasons why the IE results failed to feed into policymaking. The most important one was probably the fact that the policymakers had already acted (and changed the program) before the results of the evaluation came out. This failure is a reflection of the lack of buy-in and evaluation culture: even when a good and rigorous evaluation is allowed, many policy makers, at a deep level, fail to understand its purpose and its potential role.

From our analysis of stakeholders' testimonies we also infer that probably the political pressure from the SENA was another important cause for program closure. The IE showed strong evidence in favor of maintaining the original design of the program, with the private intermediations of ECAPs. However, this policy recommendation was not welcomed by the SENA managers, who were interested in maintaining the status quo of training policies in Colombia without private intermediation.

In contrast to the wide dissemination of Familias en Acción IE, Jovenes en Acción's IE is far less well known within government and local academic circles. The meetings for presenting the results to the SENA were regularly attended only by middle-ranking officials of the entity, while Familias en Acción's IE results were shown to the President of Colombia. The high level officials did not attend the presentations and did not welcome the evaluation's results. Since the original champions of the program had left the government and SENA enjoyed particularly strong support from the President, the discussion was avoided. In contrast, the Familias en Acción evaluation was disseminated through political and academic

seminars, news conferences, and several meetings among high-ranking government officials.

Perhaps the results from the Jóvenes en Acción IE came late and were not seen as useful by the original managers because they were unable to protect the program from institutional transition, while decisions in terms of the program's design were determined by the new operational context. While the originally conceived program disappeared, the model was adapted and transferred to the Dominican Republic with the support of the development banks, and later evaluated with a randomized design.

C. The impact evaluation of the Hogares Comunitarios de Bienestar (HCB) nursery program

As mentioned earlier, one of the major effects of the success of the Familias en Acción IE was showing the potential of a good evaluation in the public sector in Colombia. One example was the decision to implement the IE of the nursery program Hogares Comunitarios de Bienestar (HCB). HCB is the primary public program for early childhood development in Colombia. The program has operated for more than two decades, with over 61,000 nurseries in the country, serving around 780,000 children with a total annual budget of approximately US\$ 220 million.

By the time Familias en Acción was launched, program managers and government officials began to discuss whether FeA would be a good substitute for HCB or whether FeA would make HCB unnecessary. Although the designs of FeA and HCB were very different, both were interventions directed at poor children, and both aimed to support poor single mothers in need of childcare services.

Given that the Familias en Acción IE included data collection, and the implementation was about to start, evaluators thought it would be a unique opportunity to include questions regarding HCB in the FeA IE questionnaires, and conduct an empirical analysis of HCB's effects on children's nutritional status, school achievement and female labor supply. As a result, the consortium IFS-Econometría-SEI, which conducted the first IE of Hogares Comunitarios de Bienestar found large and positive impacts among HCB beneficiaries (Attanasio et al, 2006). This evaluation, which later became the topic of a paper (see Attanasio and Vera-Hernandez, 2007) was based on an Instrumental Variable approach. This technique was used because the program was effectively present in all municipalities in Colombia. It was therefore difficult to identify a control group, defined as a group of poor children who had no access to an HCB. Attanasio and Vera-Hernandez, therefore, decided to use the distance from the nearest HCB as an instrument that could affect the decision to attend but, potentially, not directly affect the outcomes of interest. This strategy is particularly appropriate for rural areas where the distances from the nearest HCB can be substantial. Later, Attanasio, Di Maro and Vera-Hernandez extended this approach using data representative at the national level (the DHS) and using as 'instruments' for the availability of HCB, the average fees at the municipality level and the number of available places over the number of potential beneficiaries. Remarkably, the results obtained with the different instruments are extremely consistent and indicated an impact of almost a standard deviation in the z-scores for height per age.

The difficulties experienced by the IFS-Econometría-SEI consortium to identify a pure 'control' group would later become the main challenge for the impact evaluation conducted by the consortium Universidad de los Andes (UA)-Profamilia: the high coverage of HCB (almost universal), and the corresponding difficulty for finding an adequate control group.

By the time the UA-Profamilia consortium won the international tendered process, the program had reached 16 years of operation, and was expanded all over the country. The vast coverage made HCB an especially difficult program for a rigorous impact evaluation. We will come back to this point later in the section.

Since its creation, Hogares Comunitarios de Bienestar has been run by the Instituto Colombiano de Bienestar Familiar (ICBF), a governmental entity responsible for implementing family policies in Colombia. HCB consists of childcare services offered through a communitarian participation scheme. When it began, the ICBF identified poor neighbourhoods and encouraged parents with children under 6 to form parents' associations. The program currently works under the umbrella of this scheme. Parents within the association elect a female called *madre comunitaria*, who is responsible for running the nursery in her house¹¹. Although the majority of program funding comes from the ICBF budget and is transferred directly to the parents' associations¹², each family additionally pays a small monthly fee that helps support a salary for the *madre comunitaria*¹³. Children who attend a HCB are entitled to childcare services, nutritional supplementation, and instruction to enhance psychosocial and cognitive development.

The HCB IE conducted by the consortium UA-Profamilia began in June 2006 with the signing of the contract. Although evaluators tried to start the field work in the same year, it was only in February 2007 that it actually began¹⁴. The methodology laid out in the Consortium's awarded proposal anticipated the use of three control groups. The first group consisted of children under 6 who live close to a HCB included in the survey, are eligible for one, do not participate in Familias en Acción and do not participate in HCB. The second was supposed to be made up of children who do not attend a HCB in communities where the program is not available¹⁵. A third group was later defined in the evaluation, consisting of treatment children with different durations within the program (intensity design)¹⁶.

Even though the evaluators had conducted pilots, and had the perception that all control groups were feasible, the real field work confirmed that the program was so widely spread that the second control group was virtually empty. Most municipalities in Colombia do have a HCB. This created tremendous challenges for evaluators, who finally relied on the first and third control group for conducting the IE. Given that the IE was based on the feasibility of the control groups this resulted in a lot of reservations about the soundness of results.

The first and third control groups were likely to have problems of selection of unobservables into the program, which were not possible to correct with matching techniques on observables or double differencing given the lack of baselines. But in addition to the

¹¹ It is the reason why the *madre comunitaria* must have a big house. Another requisite for being a *madre comunitaria* is completion of basic education, which in Colombia is equivalent to nine years of instruction.

¹² Some have argued that the involvement of parents' associations and the *madres comunitarias* have granted the program a large social base creating political inflexibilities on many decisions regarding administration and budget.

¹³ More information about the program is presented in the Annex (table A-1).

¹⁴ This is a common feature of IE in Colombia, which is mainly explained by delays in administrative procedures that usually are out of evaluators' control.

¹⁵ These neighborhoods would have the same socioeconomic characteristics of treatment children neighborhoods to be comparable.

¹⁶ Children that were early enrolled would be the treatment group and recently enrolled children would make the control group.

critiques on the feasibility of the control groups, another technical limitation of the evaluation was that the sampling scheme implemented in the program did not allow for an unbiased estimation of participation that would have been needed to use instrumental variable techniques in order to address the endogeneity of program participation. Instead of conducting a comprehensive survey of geographical areas, the consortium implemented a choice based sample approach that identified a representative sample of HCB and interviewed eligible children attending it. This implied that the survey was muted about coverage of the program, either in the localities included in the survey or in general, imposing problems on matching and instrumental variable techniques.

The alternatives given to the consortium were to derive participation using complementary data from another survey (called Sisben) and carrying out matching on the basis of odd ratios rather than propensity scores. These corrections were carried out, and some instruments were explored, although with the usual arguments for their validity and independence from the outcomes explored.

The choice of the sample approach seemed a reasonable response to the need of having representativeness at the subnational level, a feature demanded by the program managers who were interested in knowing how the different subnational administrators performed. The final sample size was over 26,000 children. The choice based sample was reasonable in that it implies lower collection costs and is commonly used in evaluations, but has implications over the potential methodologies used.

In conclusion, the feedback from peer reviews proved very useful for the final version of the evaluation and the consortium managed to include the necessary methodological corrections in the final version of the evaluation.

The HCB impact evaluation conducted by UA-Profamilia did not find effects on the nutritional status of children under 24 months and children above 48 months. However, for those aged between 25 and 48 months and exposure over 2 months the IE showed a positive and significant effect. An interesting result is a larger positive effect when the HCB's floor is made of concrete or ceramic (hard floor). The evaluation also showed positive impacts on health practices, especially among those who have been in the program for more than 16 months. As for the impacts on cognitive and psychosocial development outcomes (carried out in a sub sample), the IE showed positive effects on socio-emotional behaviors and peer interaction, and positive impacts on cognitive development of children who had been in the program for more than 16 months. In general, the results were less consistent than expected across the different group/exposure combinations.

Given the problems with the viability of the control groups, and the challenges of the nature of the HCB program, the consortium UA-Profamilia was not very successful at conducting an unbiased and rigorous IE of HCB. However, IE managers and the ICBF officials agreed that the findings from the descriptive analysis were very useful in terms of policymaking. In particular, these included findings regarding compliance of guidelines established by the program's administrators, and regional differences in their application. According to testimonies from different stakeholders, the evidence regarding some mistakes in hygiene practices and food preparation resulted in the launching of a strategy to train the madres comunitarias in these areas.

Another interesting experience from this evaluation was that managers were able to decide against a follow up survey on the basis of the technical peer-reviewing that was conducted. After discussions with the consortium and international IE experts who reviewed the technical challenges, the IE managers decided to end the evaluation at that point.

Remarkably, despite the technical difficulties during the implementation of the evaluation, this evaluation can be considered a best practice example in terms of dissemination of results. Top-level officials in the agency and technical staff interacted closely regarding the implementation of the study and built upon the findings and recommendations. A fruitful interaction between academics and policymakers developed, and this interaction led to additional research work which is enabling major positive changes in the implementation of child care policies in Colombia (See Box 2).

Box 2. Tracing the influence of evaluations: *Hogares Comunitarios*' successful dissemination and resulting interaction between academics and policymakers.

A complete dissemination plan was defined since the beginning of the HCB IE. After several rounds with technical officers and peer-reviewing by experts, the principal investigator presented the results of the evaluation in a meeting with top-level authorities including the director of the agency in charge of the program (ICBF), the Social Protection Minister, the Presidency Advisory Minister, and the Planning Minister, among others.

The numerous presentations and interactions generated buy-in from the agency's director who in turn supported dissemination of the results within the agency to all staff involved in the program at sub-national levels, including presentations of results by the principal investigator during large convention events of the mothers running the nurseries (*madres comunitarias*). For the first time the final report of the evaluation by DEPP (DNP, 2009) not only included the technical findings and summary of the evaluation but also devoted a whole chapter to the response, commitments and ensuing actions by ICBF regarding the program.

In addition, the ICBF launched concrete actions to address the weaknesses found by the IE. It conducted focus groups to better understand the descriptive quantitative findings; it revised nutrition protocols, started education campaigns in hygiene and childcare procedures and incorporated some incentives for the *madres*, among others. It also launched a series of initiatives related to the findings of the program; an evaluation of the housing improvement subsidy and the implementation of a new technical training program in child development and care for the *madres*. The evaluation had found that the *madres* had on average low education levels and lacked any training for childcare (DNP, 2009).

Finally, the principal investigator that conducted the HCB IE continued working closely with the ICBF. A prime example of this joint work is the evaluation of the technical training program for *madres* that was implemented using a randomised design (Bernal, 2010). The evaluation assesses the effects of the training program on the quality of care offered at the nurseries and the effects on participant children's nutritional and health status, cognitive and non-cognitive development. The results found that "(1) the quality of care significantly increased, (2) the pedagogical process has improved, in particular, productive activities with children and use of pedagogical resources, and (3) interaction with parents has improved. As a result of these changes, there have been positive and significant effects on participant children in treated *hogares comunitarios* on health, cognitive and non-cognitive outcomes, particularly for children between the ages of 6 months and 3 years of age" (Bernal, 2010).

D. The evaluation of the Sistema General de Participaciones (SGP) fiscal transfers' policy

One of the agreements of the new constitution of 1991 that resulted from the peace agreements with one of the major rebel groups of Colombia, was the need to strengthen the decentralization of the country, including the definition of new rules for the transfer of resources from the central level to the territorial entities. It was agreed that the fiscal transfers to territorial entities will grow at the rate of growth of the national current revenues. However, after the crisis of the late nineties, it became evident that this pace of growth was fiscally unsustainable and created serious budget inflexibilities. In order to address these challenges, the government introduced in 2001 the Sistema General de Participaciones (SGP), as one of the main reforms to tackle the financial crisis of Colombian municipalities. Before this, the indexation of transfers to the current revenues enabled the expansion of budget constraints during economic booms, but also later on generated strong inflexibilities in public finances, in particular expenditures in teacher hiring and school construction. As one may expect, this scheme also introduced high uncertainty in program coverage and resulted in low quality service delivery.

The main reforms introduced in 2001 through the SGP, were the creation of earmarked funds for each social subsector,¹⁷ the introduction of a set of incentives for management of resources both among and within social sectors, and the change of the fiscal transfer's growing rate (FTGR) during a transition period of eight years (2002-2008). The new FTGR was a fixed rate indexed to the inflation rate, plus a couple of percentage points. Although government officials wanted to establish the new FTGR for an indefinite period, the Colombian senators and representatives pushed the approval for only eight transition years arguing that this change was motivated by the economic crisis and that it would be overcome by 2008.

The SGP was launched in 2002 with high expectations regarding the changes introduced and the efficacy of the reform. Although there were some prior assessments about the decentralization process, those studies had focused on macro level variables and very little was known about the micro level performance. This knowledge gap and the fact that the Sistema General de Participaciones was an instrumental tool for Colombian social policy motivated the government to launch an evaluation of the SGP. Given the nature of the intervention, the first decision regarding the potential evaluation was that the methodology should be comprehensive and suited to the nature of the transfers system and not rely exclusively on impact evaluation techniques. In this way, Public Expenditure Tracking Surveys (PETS) was the suggested methodology in the terms of reference (TOR) of the study.

Having defined the evaluation approach, in 2005 the DEPP presented the first draft of the terms of reference (TOR) and began a wide consultation process with all potential stakeholders, a standard practice for DEPP. However, unlike prior evaluations, the SGP system involved many actors, entities and political interests at all government levels, which made the consultation process very complex and introduced delays in the implementation of the study. The final version of the TOR was ready only by the end of 2006 and still did not manage to elicit strong support from all SGP stakeholders¹⁸. According to interviews for the

¹⁷ Education and health within the reform of 2001, and water in the constitutional reform 2007.

¹⁸ For instance, the stakeholders from the health sector had expectations of an evaluation of the entire health sector reform, rather than an evaluation of the sector within the context of the *Sistema General de Participaciones*.

present report, the inconformity among some stakeholders would reinforce the main challenge of this evaluation, the reluctance from SGP managers at various government levels and the weak buy-in of the study.

Another factor that contributed to the delay of the evaluation's implementation was the difficulty of conflating funds from different sources to finance the evaluation¹⁹. The lacking of resources to finance the study in its entirety meant that funding came from a variety of sources, including funds from development banks as well as funds from the budget of the Ministry of Social Protection. It was challenging to set up an administrative arrangement for resources governed under different rules. After several consultations among the DEPP and development banks, the evaluation managers decided that the feasible way to proceed was to use the United Nations Development Program (UNDP) as a fund administrator, conflating the resources there. By the end of 2006, once the DEPP had obtained approval of the terms of reference and the identification of the administrator, the international tendered process was launched.

The evaluation was awarded to a consortium consisting of the Centro Nacional de Consultoría (Colombian consulting firm) and Econosul (Argentinean consulting firm), CNC-Econosul. Although the firms were informed of the decision in mid April 2007, there was a need to request a waiver to the development banks before the negotiation stage. This administrative procedure, provided in banks' operational manuals, requires some time and the contract was signed only in December 2007. Somewhat unfortunately, the study officially began the day after the approval of a new reform to the Sistema General de Participaciones. Hence, the launch of the evaluation revived the debate regarding the timeliness of the study and the usefulness of its future findings. Like in the past, this unconformity was arguably one of the reasons for the weak interest from SGP administrators. Without doubt, their reluctance hampered the smooth development of the study during the two years of its implementation.

The consortium CNC-Econosul proposed an evaluation comprising four components. The first, an evaluation of the institutional capacity and the service delivery process, which had two main goals: to analyze the overall conditions for social services delivery, and to quantify potential leakages in the fiscal transfers' flows. The purpose of the second component was the evaluation of SGP's outputs. It intended to assess the relationship between public expenditure and variables such as coverage, quality, efficiency and targeting of poor households. The third component was an impact evaluation of the fiscal transfers. It aimed at assessing the effect of SGP at two levels of analysis, municipalities and households. Finally, the fourth component sought to evaluate the SGP policy within the context of the Colombian decentralization policy, contextualizing the results and proving a framework to assess its relevance.

In order to achieve the main evaluation goals, the consortium applied the Public Expenditure Tracking Survey (PETS) and Quantitative Service Delivery Survey (QSDS) methods. A PETS studies the flow of public funds among different levels of government (Reinikka & Smith, 2004) while a QSDS focuses on the characterization of service delivery exclusively. Although the terms of reference did not require the application of QSDS, the consortium proposed the method in order to obtain a complete picture of the SGP implementation. Once the first version of the methodology report was submitted, evaluators began the survey and field work design. Because PETS tracks flow of resources among different levels of government and compares information reported from different

¹⁹ The research cost US\$1.2 million.

stakeholders, one of the first tasks was to gather administrative data from ministries and government entities at the national level.

Even though that could have been thought as an easy part of the process, evaluators and evaluation managers indicated during interviews that this was actually one of the main challenges of the study. With limited interest in the evaluation, some government officials were reluctant to share the administrative data and the DEPP was unable to enforce prompt collaboration. The process was so frustrating that at some point the consortium decided to modify the questionnaires and include information that could be found via administrative data, but was not being made available to the evaluators. Likewise, although questionnaires were drafted on time, the field work could not begin on the date it was originally planned. As we have mentioned, unlike prior evaluations managed by the DEPP this one involved many program managers – some of them also within the DNP- and this made it difficult to obtain approval of the questionnaires within the time originally allotted. To sum up, it took ten months to gather the administrative data and eight months to finalize survey design.

The high complexity of the evaluation was also reflected in the field work. Undoubtedly, one factor that contributed to exacerbate complexity was the scope of the study. Although the majority of PETS conducted in the world were based on small samples and not necessarily statistically representative, the terms of reference required a sample statistically representative within 7 categories, and therefore, it ended up having a size of 223 municipalities, 32 departments, 620 schools, 360 hospitals and 269 providers of water and sewage services. In the original schedule the field work was meant to be done within a period of two months, and it finally took 8 months to get a clean data base for the quantification of leakages and the analysis of service delivery.

Once the consortium overcame obstacles for conducting the field work and the data base was ready for the analysis, the overall evaluation became an easier goal to accomplish. Early findings were presented in mid August 2009, and a final report was submitted in mid November 2009. The consortium presented the results to the DEPP and the other directorates involved in SGP implementation within the DNP and got some feedback from program managers. Dissemination efforts were scarce and to date it is unclear if there have been discussions on concrete measures and steps derived from the study recommendations within the government. To some extent this could be a consequence of the weak buy-in from program managers and an expected reaction to the adverse evaluation results. The consortium found leakages of SGP fiscal transfers²⁰ within all sectors analyzed, problems of low institutional capacity among municipalities and services providers, and therefore, no impact of transfers on the quality of the services delivered. Box 3 summarizes major findings from this evaluation.

As the literature highlights, the main obstacle for managing a PETS is the resistance among stakeholders to be made accountable²¹. In sum, implementation of the evaluation proved particularly challenging for the consortium. On one hand, there was evident reluctance from program managers and fears of being pointed out for potential performance issues –which could hypothetically eventually trigger legal responsibilities. On the other hand, the evaluation was very complex in terms of the number of actors and stakeholders involved the intended large scope, and the limited authority and resources of DEPP to fully support it.

²⁰ Education, health and water and sewerage services.

²¹ *Ibidem*

Box 3. Major findings from the SGP fiscal transfers' policy evaluation²².

SGP's evaluation reveals that poor institutional capacity has been an important constraint to the provision of adequate social services in some municipalities. Even though results show evidence of progress in more developed areas, on average, social policy could do more to enhance service delivery in low income towns. This evaluation also shows evidence of leakages associated with transfers of resources from different government levels to service providers. These leakages were identified across all sectors and account for a non-negligible percentage of the amount of fiscal transfers. Finally, the results indicate that there is an important increase in services coverage that can be attributed to SGP. However, such an increase does not necessarily reflect more efficient and high quality services.

Education Sector

Constraints to delivering high quality services in this sector mainly come from low institutional capacity of some public schools and the agencies in charge of running sector policies at the local level, the so called education secretaries. This evaluation also found leakages due to teacher absenteeism. Results indicate that 8.2% of fiscal transfers from central government to education certified municipalities²³ and 7.3% of fiscal transfers from central government to non-certified municipalities are lost due to this phenomenon. Given that 96% of the education resources go to teacher wages, this is one major concern of different stakeholders across this sector. Other leakages include overpricing of procured items supplied for schools, and expenditure of earmarked transfers outside this sector. Even though the impact evaluation shows an increase in coverage attributable to SGP, findings also indicate that more fiscal transfers do not necessarily enhance quality and efficiency of education services. In fact, an increase in transfers has a negative impact on the average score of the ICFES test²⁴ at the local level.

Health Sector

Major concerns in delivering health care services are also related to the capacity to fulfill mandates dispersed among different stakeholders. For instance, 62% of health certified municipalities and 48% of non certified municipalities had delays in payments to insurance companies, denominated EPSs, (entidades promotoras de salud). In the executive summary, evaluators indicate that this problem is mostly due to operative obstacles behind the identification of population insured under the subsidized health care scheme. In particular, the evaluation indicates that, in 2008, 42.5% of the newly insured have not received the ID that grants access to the health services. Given that fiscal transfers in this sector are mainly determined by the insured population, this problem also represents one major leakage in this sector. On the other hand, the impact evaluation showed an increased in coverage attributable to SGP as well as a positive and significant effect on prevention of health issues.

Water and sewage services

Adequate service delivery in this sector has also been hampered by institutional capacity issues. Results from this evaluation indicate that only 66% of municipalities have an appointed person in charge of running and monitoring water and sewage service delivery. Likewise, descriptive statistics shows that compliance to mandatory rules in the sector is very low. Only 57% of municipalities actually apply methodologies enacted by the national

²² This summary is based on Union Temporal Centro Nacional de Consultoría-Econosul (2010).

²³ Generally speaking, being certified in education means that these municipalities have more autonomy for running education policies than those who do not hold this certification.

²⁴ Standardized test, similar to the American SAT.

Box 3. Major findings from the SGP fiscal transfers' policy evaluation (continued)

regulatory agency. In addition, major leakages in this sector include: (i) differences between the amount of pesos that central government reported as transferred to local government and the amount that municipalities include in their budgets and spend in this sector; (ii) expenses of earmarked transfers outside the sector (ranges from 9.9% in small municipalities to 3.8% in big cities); and (iii) allocation of subsidies of higher value than the actual local deficit.

4. Conclusions and key lessons

The experiences presented in the former chapter illustrate some of the challenges that evaluation practitioners face.

The complex nature of IE projects, the effect that evaluations have in shaping public policies and the results on interventions that directly affect the welfare of communities and families, create particular challenges for each evaluation. Despite this, our qualitative research shows some commonalities among the cases analyzed and some standards that tend to produce an "Ideal IE", defined as one that is technically robust, produces useful findings and is politically accepted and taken seriously by decision makers.

In the first section, we defined 11 standards that should be considered for the development of an ideal IE. In this section, we briefly present an assessment of the level of adoption of each standard in the four cases discussed, which is summarized in table 4.1. The analysis was based on the desk review of documentation from the M&E system and the structured interviews conducted for the study. At the end of this chapter we draw some lessons learned from these experiences and offer recommendations for IE practitioners.

On the General Considerations

While revisiting the set of general considerations, we found that the principle of independence was in general met by the practice of contracting evaluators externally. However, sometimes problems arose when evaluation managers and program managers were peers like in the case of the SGP. Unlike other experiences, this was the first time that the DEPP was undertaking an evaluation of policies managed by other peer DNP directorates. Somehow this situation limited the full independence of the evaluation process.

In terms of capacity building, the interviewed stakeholders particularly highlighted the case of Familias en Acción. The evaluators saw the absence of IE skills in Colombia as an opportunity for transferring capacities and learning about the institutional arrangements of social policies in Colombia, resulting in a win-win situation.²⁵

Regarding quality control, another important observation is that all evaluations benefited from the double academic-evaluator role of their authors. The JeA and FeA evaluations underwent additional academic screening and met high quality standards, even if a well defined process of quality control was not planned in advance. Both IEs benefited from the academic reviewing of the published technical papers.

In addition, peer-reviewing proved extremely important for quality control. HCB benefited from a through peer-reviewing process, which was defined and planned since the beginning of the IE. In this case, as in the case of the SGP, the World Bank and Inter-American

²⁵ "It was the first time that an impact evaluation was conducted in Colombia, and all stakeholders, including evaluators, learned from the process" testimony from program manager, June 15th, 2010.

Development Bank facilitated top-quality international peer-reviewers, some of whom were Bank staff while others were externally contracted. Peer reviewers could be local experts hired such as professors, development bank's experts, and experienced public servants, among others.

Finally, although technical rigor is essential, buy-in also proved necessary for a successful evaluation process. The interviewed stakeholders agreed with the statement that the Jóvenes en Acción IE is the most rigorous IE that has been undertaken in Colombia. Unfortunately, the diverging interests and the reluctance from program managers also resulted in low contribution to capacity building, little interaction among stakeholders and poor dissemination of evaluation findings. Paradoxically, JeA is more cited within the international academia than in Colombia.

On the conditions before starting an IE

Once the case for an evaluation is clear and the decision to conduct it has been made, the first step is to identify the key issues and evaluation questions that it aims to address. It is clear that the process of identification of the evaluation questions is a critical step to later ensure the relevance of the findings. Sinergia manages its evaluations under a multiyear program or agenda, and all the evaluations analyzed here were included in the program as important evaluations. In general, in all cases the rationale for the evaluations was clear and important questions were identified in the terms of reference (TOR).

However, more work could still have been done in explicitly laying out the assumptions and hypothesis underlying each of the program/policies. For instance, the HCB evaluation did not request an analysis of time and labor market gains of children's mothers and primary caretakers. In spite of being a fundamental reason for existence of early childcare programs, this was not included within the scope of the TOR.

Both SGP and HCB were retrospective impact evaluations in the sense that the programs or policy were already in place and functioning when the evaluation was designed. In contrast, Familias en Acción and Jóvenes en Acción were prospective, and showed a narrower object and scope definition. In both cases, the IE was designed almost simultaneously with the programs, and the intervention followed existing models (Progesa-Oportunidades) and was led by the principal investigator jointly with a unique agency. Perhaps this helped to fulfil more easily the standards of relevance and clarity in the TOR. Despite not being timely, the SGP evaluation could also stand out as a good example in identification of relevant issues and evaluation questions in the TOR – this despite the fact that it had an overarching scope.

A second step is to decide the best methodological approach to address the evaluation questions. If the approach chosen is an impact evaluation, then one has to assess if the impact evaluation is feasible, and then define a realistic scope or the need to use complementary methods. For instance, SGP correctly identified that the best suited methodology was not an IE but PETS. In the case of HCB, the existence of valid comparison groups challenged the IE viability, and this was not anticipated correctly.

In some cases, the demand for sub-national and sectoral representativeness imposed very ambitious scopes that had consequences for the sampling approach chosen by the HCB consortium and by the implementation of PETS in SGP. The extensive consultation processes, in particular when evaluation and research capacities of stakeholders are limited, revealed important trade-offs in the case of SGP. While it is often politically unfeasible to impose one narrow technical approach, trying to please all actors involved also comes at a

cost. These difficulties should be more manageable when the agency commissioning evaluations enjoys top level ranking and high powers.

Undoubtedly there was room for improvement in the evaluation's TOR, and this was an aspect repeatedly mentioned by the interviewees. Also further TOR standardization could help, although there are always a number of features specific to each program and context. Perhaps availability of thematic reviews in key areas which help to summarize existing knowledge gaps, experiences from prior evaluations, and studies in other contexts could facilitate the definition of evaluation questions and increase the relevance and clarity of local TOR.

On critical aspects during IE development

In general, all evaluations were able to respond correctly to the evaluation questions posed in the TOR, and presented recommendations well supported by the findings and methodologies used.

Undoubtedly the FeA and HCB cases stand out as good examples of the positive interaction that emerged from the continuous communication with stakeholders during the evaluation. Evaluators cultivated managers' trust and managers perceived an opportunity to access high quality technical feedback and improve their programs through the researchers. Unlike other cases, FeA and HCB eventually developed strong support from the program managers, overcoming the initial resistance. Both program managers and evaluators highlighted the evolution of support for these evaluations.

From what we learned of HCB stakeholders' testimonies, the key element that enabled this successful combination was the strong support from the former agency director who launched the evaluation²⁶ and the fruitful dialogue established between program managers and evaluators during the study implementation. The evaluation of HCB was seen as relevant and useful since the beginning, for not only establishing the impacts of the program on children's nutritional and early development situation, but also for assessing whether FeA would be a good substitute for HCB or not.

In contrast, interaction with key stakeholders was very poor in the cases of SGP and JeA. Interests at stake were very high and managers preferred to maintain status quo. Potential benefits from the evaluation were insufficient to overcome the reluctance.

From a more general perspective, the cases studied reinforce the idea that institutions matter and the context in which the evaluation is carried out determines to a large extent the political complexities that evaluators face. One disadvantage of evaluation systems inside the government and without strong political backing is that they depend excessively on buy-in from program management²⁷. This makes evaluations more vulnerable to political concerns, reluctant staff, transition of administrations, or censoring of findings. Indeed, some analysts have argued that SINERGIA requires a governance reform supporting its independence or increasing its position within the executive. For instance, an interesting arrangement that could support higher enforcement and disclosure abilities is a mixed governance board with representatives from academia, similarly fashioned like the Mexican system, Coneval.

²⁶ "We seriously believed in the importance of evaluation as a useful tool for improving the impact of public policies" Former Director of the agency that runs the HCBs.

²⁷ See Briceño and Gaarder (2009) for a discussion on trade-offs of various type of institutional arrangements.

Table 4.1. Quality standards applied to the implementation of four Colombian leading impact evaluations.

Standards		Program/Policy			
		FeA	JeA	HCB	SGP
General Considerations	Independence of the evaluation process	●	●	●	◐
	Contribution to capacity building and promotion of evaluation culture	●	◐	◐	◐
	Quality control	●	●	◐	◐
	Buy-in from stakeholders	●	◐	●	◐
Before starting the IE	Clarity of the evaluation's object and scope, evaluation questions	●	●	◐	◐
	Feasibility of the evaluation	●	●	◐	◐
During implementation of the IE	Active interaction among stakeholders	●	◐	●	◐
	Strong relationship between findings and recommendations	●	●	●	●
	Answers to evaluation questions in the TOR	●	●	●	●
After implementing the IE	Quality dissemination	●	◐	●	◐
	Timeliness and utilization of the evaluation findings	●	◐	●	◐
Overall Assessment		●	◐	◐	◐
Key: High ● Medium ◐ Low ◐					

On the conditions after completion of the IE

What happens after the results come out shows that solely achieving a rigorous and unbiased impact evaluation does not guarantee the utilization of evaluation findings and the full exploitation of the investment in the study.

The best example of an evaluation process that does not conclude with the publication of a report is the case of HCB. Although the study was initially criticized for the technical difficulties, it fully exploited the evaluation findings, especially those shown by the descriptive statistics and qualitative work.

In the end, the negative findings were seen as an opportunity to improve and not just critique the performance of government officials. Moreover, the ICBF launched different measures to correct the problems detected by the IE.

The dissemination events and follow-up plans were carefully designed and carried out. For the first time in Sinergia's existence the evaluation report incorporated a chapter laying out the response from the ICBF agency and a proposal with key measures that the agency wanted to voluntarily adopt in order to address the weaknesses identified by the evaluation (see DNP, 2009).

The cases studied also showed that presentation of early findings proved useful to elicit support and maintain interest in the IEs. The case of FeA showed that early findings presented to the President, academia, civil society, and the press enabled buy-in for the later stages of the evaluation and relieved time pressures. In addition, FeA databases were published in DNP's website, and have been extensively used in follow-up work by students, researchers and ministry officers.²⁸

In contrast, so far, there has been a very incipient dissemination of the evaluation of the SGP fiscal transfers' policy. No report has been made public and high level discussions were delayed. This is unfortunate given the relevance of the topic, the tremendous effort that it demanded from both evaluation managers and evaluators, and the amount of resources invested in its implementation.

Five Key Lessons

We end with five key lessons for practitioners.

1. *Invest in the preparation of good terms of reference (TOR).* The relevance of the study depends critically on the correct identification of the evaluation questions. Once the case for an evaluation is clear and the decision to conduct it has been made, identify the key issues and evaluation questions that it aims to address. Use theme reviews summarizing knowledge gaps, prior evaluations and studies in other contexts, which could be helpful in correctly defining the evaluation questions and increasing the relevance and clarity of local TOR.
2. *Decide on the best methodological approach to address the evaluation questions.* If the approach chosen is an impact evaluation, consider if the impact evaluation is feasible (viability of counterfactual scenario), and then define a realistic scope and the need to use complementary methods. One should prepare contingent plans while attempting to implement a randomization approach.
3. *Ensure evaluation quality.* Both academic screening and experts' peer-reviews should be used to enhance rigor and technical quality of the evaluations. Identify adequate peer-reviewers (local experts such as professors, development banks' experts, and experienced public servants, among others) and plan for a thorough peer-review process.
4. *To foster evaluation ownership, lay out the incentives for dialogue between the involved parties.* Make sure that evaluators know that a fruitful interaction with managers can result in future research opportunities; and that managers realize that researchers can provide them with high quality technical feedback to improve their programs and show better results.
5. *Conduct quality dissemination.* The evaluation process does not end with a published report. Prepare a clear dissemination plan and engage all parties along the evaluation process: from presentations of the methodological approach at the beginning of the evaluation, presentations of early findings to elicit support and maintain interest, to

28 See Garcia and Hill (2009) and Baez and Camacho (2011), among others.

presentations and discussions of draft final reports with technical officials. Predefine a top-level target audience and venue to present the evaluation recommendations. Media is potentially a powerful disseminator of evaluation's findings. Finally, document the stakeholders' responses, the discussions and recommendations adopted, and the commitments made on account of the evaluation.

Annexe

Table A-1. Context for the implementation of the main IE conducted in Colombia²⁹.

Feature	Program/Policy			
	FeA	JeA	HCB	SGP
Goals	Protect human capital of poorest children through the following goals: <ul style="list-style-type: none"> • Increasing school attendance at primary and secondary levels. • Improving health care for children under 7 years old. • Complementing family income with the purpose of increasing food and education expenditures. 	Reduce youth unemployment rate among poorest population.	Protect poorest children through the following goals: <ul style="list-style-type: none"> • Increasing of child care 	Enhance decentralization performance through the accomplishment of the following goals: <ul style="list-style-type: none"> • Increasing coverage of the social services of education, health and water. • Augmenting quality of social service delivery. • Improving targeting of social public expenditure. • Improving efficiency of social public expenditure.
Type of intervention	Conditional Cash Transfers (CCTs) to participant households with children under 17 years old	CCTs and training to urban youth unemployment.	Child care to urban children under 7 years old.	Funds allocated to municipalities, departments and districts to implement education, health and water policies.
Components	Two: CCT for children under 7 who attend	Two: Three months of class in-classroom training	Multiple: child care, nutritional	Multiple

²⁹ The structure of this table is based on Moreno et al (2009), pp 14-17.

	to medical checks; CCT for children from 7 to 17 who attend to school.	and three months of on the job training.	supplementation, cognitive and psychosocial development activities.	
Feature	Program/Policy			
	FeA	JeA	HCB	SGP
Target population	Households in extreme poverty with children aged 0 to 17 placed in the two lowest deciles of the income distribution.	Youth between the ages of 18 and 25, who were unemployed and who were placed in the two lowest deciles of the income distribution.	Households in poverty with children aged 0 to 6.	All municipalities from the country.
Program duration	Open ended	2001-2005	Open ended	Open ended
Timing of intervention	All elements concurrently implemented	All elements concurrently implemented	All elements concurrently implemented	All elements concurrently implemented
Institution responsible for implementing the program	Acción Social	Acción Social	Instituto Colombiano de Bienestar Familiar (ICBF)	<ul style="list-style-type: none"> • National Planning Department • Ministry of Education (ME) • Ministry of Social Protection (MSP) • Ministry of Housing and Local Development
How the program fits in the country's social policy	Began as a temporal strategy for protecting poor children from the economic downturn of 1999. Currently, it is the pillar of the most important social policy.	Temporal strategy for protecting poor youth from the economic downturn of 1999.	One of the main social programs directed to poor children in Colombia.	The financial and normative instrument for the implementation of the main social policies in Colombia.

Type of funding	Loan	Loan	Loan	National Public Budget
Cost of Evaluation	Aprox. USD 2 million	Aprox. USD 0.7 million	Aprox. USD 1 million	Aprox. USD 1,25 million
Funding source	World Bank/Inter-American Bank	World Bank/Inter-American Bank	Colombian Government	Colombian Government
Feature	Program/Policy			
	FeA	JeA	HCB	SGP
Impact evaluation required by funding agency	Yes	Yes	No	No
Set-aside funds for impact evaluation	Yes	Yes	Yes	No (but ME and MSP)
Legal requirements for the evaluation	No	No	No	No
Organization responsible for the evaluation	Independent, selected by Colombian government under multilateral agencies procurement rules.	Independent, selected by Colombian government under multilateral agencies procurement rules.	Independent, selected by Colombian government under Colombian law.	Independent, selected by Colombian government under multilateral agencies procurement rules.
Hiring organization	United Nations Development Program	United Nations Development Program	United Nations Development Program	United Nations Development Program
Procurement process	Competitive procurement process	Competitive procurement process	Competitive procurement process	Competitive procurement process
To whom evaluator is accountable	National Planning Department	National Planning Department	National Planning Department	National Planning Department
Time period of evaluation	Three years	18 months	20 months	Two years

References

Attanasio O., (2009a). "Evaluating a Conditional Cash Transfer in Colombia. Evaluation, politics and all that" Presentation at the seminar "Evaluation des politiques de développement" College de France, June.

Attanasio O., A. Kugler and C. Meghir. (2009b). "Subsidizing Vocational Training for Disadvantaged Youth in Developing Countries: Evidence from a Randomized Trial" IZA Discussion Paper No. 4251.

Attanasio O., and M. Vera-Hernández. (2006). "Medium and long run effects of nutrition and child care: Evaluation of a community nursery programme in rural Colombia" The Institute for Fiscal Studies, EWP04/06.

Baez, Javier and A. Camacho (2011) Assessing the Long-term Effects of Conditional Cash Transfers on Human Capital. Evidence from Colombia. The World Bank. Policy Research Working Paper 5681. Independent Evaluation Group.

Bernal, R. (2010). "The Impact of a Technical Training Program for Childcare Providers on Children's Well-being".

Briceno, B. and M. Gaarder (2009). Institutionalizing Evaluation: A review of International Experience. International Initiative for Impact Evaluation. DFID.

Cuesta, L. and J. Núñez (2006). Evolución de las políticas contra la pobreza: de la previsión social a las transferencias condicionadas. , " Universidad de los Andes, CEDE Working Paper Series. 2006-31. ISSN 1657-7191

DNP. (2010). "Evolución de SINERGIA y evaluaciones en administración del Estado". Dirección de Evaluación de Políticas Públicas, Departamento Nacional de Planeación, Bogotá D.C.

DNP. (2009). "Evaluación de impacto del programa Hogares Comunitarios de Bienestar del ICBF". Dirección de Evaluación de Políticas Públicas, Departamento Nacional de Planeación, Bogotá D.C.

DNP. (2008). "Programa Familias en Acción. Impactos en capital humano y evaluación beneficio-coste del programa" en Evaluación de Políticas Públicas, No. 4. Dirección de Evaluación de Políticas Públicas, Departamento Nacional de Planeación, Bogotá D.C.

DNP. (2006). "Programa Familias en Acción. Resultados de impacto un año después de implementado el programa" en Evaluación de Políticas Públicas, No. 4. Dirección de Evaluación de Políticas Públicas, Departamento Nacional de Planeación, Bogotá D.C.

DNP (2004). "Red de Apoyo Social. Conceptualización y Evaluación de Impacto" en Evaluación de Políticas Públicas, No. 3. Dirección de Evaluación de Políticas Públicas, Departamento Nacional de Planeación, Bogotá D.C.

DNP. (2000a). "Plan Colombia. Red de Apoyo Social: Programa Manos a la Obra. Componentes: Proyectos Comunitarios y Vías para la Paz". Conpes No. 3075.

DNP. (2000b). "Plan Colombia. Red de Apoyo Social: Programa de Subsidios Condicionados y Capacitación Laboral de Jóvenes Desempleados de Bajos Recursos". Conpes No. 3075.

García, Sandra and Jennifer Hill (2009). "The Impact of Conditional Cash Transfers on Children's School Achievement: Evidence from Colombia," Universidad de los Andes, CEDE Working Paper Series. Documento 005403

Gaviria, A. and J. Núñez (2003). "Evaluating the impact of SENA on earnings and employment", *Planeación y Desarrollo*, Vol XXXIII, No. 3 y 4, Julio-Diciembre.

Medina, C. and J. Núñez (2005). The impact of public and private job training in Colombia. Washington. Research Network Working Paper # R-484, Red de Centros de Investigación. Banco Interamericano de Desarrollo.

Moreno, L. L. Campuzano, D. Levy and R. Blair. (2009). "Toward closing the evaluation gap: lessons from three recent impact evaluations of social programs in Latin America and the Caribbean". *Well-Being and Social Policy*, Vol 5, Num 2, pp. 1-23.

OECD. (2010). "Quality Standards for Development Evaluation" DAC Guidelines and Reference Series.

Reinikka, R. y N. Smith (2004), "Public Expenditure Tracking Surveys in Education". *UNESCO*. International Institute for Educational Planning.

Unión Temporal Centro Nacional de Consultoría-Econosul. (2010). "Evaluación Integral del Sistema General de Participaciones en Educación, Salud y Propósito General. Resumen Ejecutivo"

Vásquez, M. "Spot Check" or Sample-Based Site Monitoring: Application in Colombia Familias en Accion.

White, H. (2009) Theory-based Impact Evaluation: principles and practice. Working paper 3, International Initiative for Impact Evaluation, New Delhi

World Bank. (2006). "Impact Evaluation- the experience of the independent evaluation group of the World Bank". Independent Evaluation Group.

World Bank. (2009a). "Institutionalizing impact evaluation within the framework of monitoring and evaluation system". Independent Evaluation Group.

World Bank. (2009b). Insider Insights: Building a Results-Based Management and Evaluation System in Colombia. Independent Evaluation Group.

Publications in the 3ie Working Paper series

Behind the scenes: managing and conducting large scale impact evaluations in Colombia by Bertha Briceño, Laura Cuesta and Orazio Attanasio, Working Paper 14, December 2011

Can we obtain the required rigour without randomisation? by Karl Hughes and Claire Hutchings, Working Paper 13, August 2011

Sound expectations: from impact evaluations to policy change by Vanessa Weyrauch and Gala Díaz Langou, 3ie Working Paper 12, April 2011

A can of worms? Implications of rigorous impact evaluations for development agencies by Eric Roetman, 3ie Working Paper 11, March 2011

Conducting influential impact evaluations in China: the experience of the Rural Education Action Project by Mathew Boswell, Scott Rozelle, Linxiu Zhang, Chengfang Liu, Renfu Luo, Yaojiang Shi, 3ie Working Paper 10, February 2011

An introduction to the use of randomized control trials to evaluate development interventions by Howard White, 3ie Working Paper 9, February 2011

Institutionalisation of government evaluation: balancing trade-Offs by Marie Gaarder and Bertha Briceno, 3ie Working Paper 8, July 2010

Impact Evaluation and interventions to address climate change: a scoping study by Martin Prowse and Birte Snilstveit, 3ie Working Paper 7, March 2010

A checklist for the reporting of randomized control trials of social and economic policy interventions in developing countries by Ron Bose, 3ie working paper 6, January 2010

Impact evaluation in the post-disaster setting by Alison Buttenheim, 3ie Working Paper 5, December 2009

Designing impact evaluations: different perspectives, contributions from Robert Chambers, Dean Karlan, Martin Ravallion, and Patricia Rogers, 3ie Working Paper 4, July 2009. Also available in Spanish, French and Chinese

Theory-based impact evaluation by Howard White, 3ie Working Paper 3, June 2009. Also available in French and Chinese.

Better evidence for a better world edited by Mark W. Lipsey University and Eamonn Noonan, 3ie & The Campbell Collaboration, 3ie Working Paper 2, April 2009

Some reflections on current debates in impact evaluation by Howard White, 3ie Working Paper 1, April 2009

For the latest 3ie working papers visit:
http://www.3ieimpact.org/3ie_working_papers.html