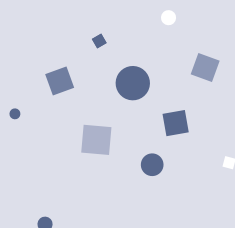


Série Notes méthodologiques

*exPost*  
*ExPost*



## Fostering impact evaluations at the Agence Française de Développement: a process of building in-house ownership and capacities

Jean-David Naudet and Jocelyne Delarue  
Agence Française de Développement



Département de la Recherche  
Division Évaluation et capitalisation  
Agence Française de Développement  
5, rue Roland Barthes 75012 Paris < France  
[www.afd.fr](http://www.afd.fr)

## **Disclaimer**

The analysis and conclusions of this document are those of the authors. They do not necessarily reflect the official position of the AFD or its partner institutions.

Director of the publication : Jean-Michel SEVERINO

Director of the redaction : Jean-David NAUDET

ISSN : Coming soon

Dépôt légal : December 2007

Layout : Solange Münzer

## TABLE OF CONTENTS

<b>Abstract</b>	<b>4</b>
<b>1. AFD's strategy towards impact evaluation</b>	<b>5</b>
1.1. The impact issue relates to knowledge production and results-based management	5
1.2. Impact evaluations are an important part of the evaluation activity	6
<b>2. A set of experiences</b>	<b>8</b>
2.1. A preliminary phase	8
2.2. The second impact assessment financed by AFD	9
2.3. A randomised controlled trial of micro-finance in Morocco	10
2.4. A randomised controlled trial of micro health insurance in Cambodia	12
<b>3. Intermediary lessons</b>	<b>15</b>
3.1. Institutional learning	15
3.2. Impact evaluation for a bilateral donor	15
3.3. Perspectives	16
<b>References</b>	<b>18</b>

## Abstract

In 2006, the Evaluation Unit of the Agence Française de Développement (AFD) was incorporated into the Research Department in order to integrate a research strand into the work of evaluation. This new approach has led AFD to undertake rigorous impact evaluations in partnership with academic teams. Two such evaluations have already been completed on Madagascar (microfinance) and Guinea (agricultural development), and two randomised control trials have recently been launched in Morocco (microfinance) and Cambodia (micro health insurance). In addition to promoting accountability, the main objective is to produce sound knowledge on development matters and to contribute to national policy-making. Moreover, AFD's and its local partners' substantial involvement in these projects is aimed at building in-house capacities. The challenge is now to address a growing number of in-house demands for rigorous impact evaluations, under time and budget constraints, using a variety of methods.

## 1. AFD's strategy towards impact evaluation

### 1.1. The impact issue relates to knowledge production and results-based management

The mandate of the Agence Française de Développement is to contribute to the funding of economic, social and/or environmental development projects.

AFD provides development assistance to the public sector (state administrations, public enterprises and local governments), the private sector and local associative networks. It offers a broad array of financial instruments to help implement sustainable development projects. AFD's goals are to (i) reduce poverty and inequalities by 2015 (MDGs), (ii) promote economic growth and (iii) protect global public goods (climate, biodiversity and global health).

Since the beginning of this decade, AFD has been engaged in the renewal of its strategic orientations, carried out within a broader process of French co-operation reform.

Amongst the strategic shifts experienced by AFD, two are worth particular mention. The first is the implementation of results-based management in line with Paris Declaration commitments. As early as 2002, the first AFD strategic plan (POS I) expressed strong concern about the impact of its operations. It recommended developing a results-oriented monitoring system partly based on impact indicators, and implementing funding selectivity grounded on impact assessments.

AFD Management gives high priority to ensuring that AFD's assistance is focused on development results and impacts. At all levels, the attention is now on increased and demonstrated effectiveness of development assistance.

The 'Management for Development Results' approach is used systematically throughout the project cycle. Aggregated indicators monitor expected and actual development results. Their definitions are standardised and harmonised with those of international agencies. Monitoring the contribution to the MDGs measures France's and AFD's commitments in terms of resources and results. The economic analysis of development projects goes beyond their financial sustainability. Economic costs and benefits are assessed for society as a whole, including environmental goods and services. Analysing how each stakeholder group benefits from a project will inform the choice of transfer mechanisms.

The second shift is towards a clear focus on knowledge production as a necessary complement to financial activity. In line with the second strategic plan (AFD, 2007), this focus will be particularly placed on major ODA-related topics in order to contribute to French policy stances, to enable participation in partners' capacity-building and to fuel international debate.

These new strategic directions have helped to raise a collective query about the impact issue. One consequence has been the growing awareness of just how little knowledge has been produced about the impacts of AFD operations (referred to as "knowledge shortage" in the "Evaluation Gap" report written by the Center for Global Development – CGD, 2006).

Although certain types of projects or financing products, such as rural roads or micro-finance, have mobilised substantial funding, AFD had until recently never conducted rigorous research on what actually works, and what is attributable to its programmes. Past evaluations have produced very little information about middle- or long-term effects, and none about the net change in outcomes attributable to the projects. Very understandably, project managers focus on project design and implementation in the early project phases, and leave until later the decisions required for preparing sound evaluations. So far, there has been very little incentive to design an appropriate project-related information system enabling ex post impact assessment, and to offset the cost and development time of such a system.

The new AFD approach of strategy-driven operations and of management for development results means that the operational services are increasingly required to demonstrate impact in order to obtain funds. They are thus placing greater emphasis on plans to capitalise on and measure the impact of their operations. For the moment, however, there is a lack both of human resources and budget to carry this out on a broad scale. It seems important that AFD first test whether rigour in impact evaluations, rather than focus on accountability or process, is able to improve the quality of feedback on operations.

At the same time, the international debate about impact evaluation reveals conceptual, methodological and practical difficulties. The attribution/contribution question comes into play when results-based management is implemented, particularly for an institution such as AFD, which is often involved in co-financing activities. The double meaning of the term “impact” in the development discourse—either long-term effects or strictly attributable effects—remains a permanent source of ambiguity. The subtle difference between impact and additionality—which is a concept more frequently used (also with much ambiguity) for financial mechanisms and particularly as regards climate change—is a concern for an institution involved both in direct financing of public policy in poor countries and in partici-

pating in financial incentives for global public goods in emerging countries. Methodological questions are directly linked with these conceptual debates: how are long-term effects to be measured and attributed, how is a counterfactual to be built, what kind of baseline is necessary, which impact indicators are relevant, etc.

This period of debate is highly enriching for AFD. It is helping to persuade AFD staff that a “one-size-fits-all” conceptual and methodological approach cannot constitute the sole answer to the shortage of knowledge on impacts, and that AFD cannot gain much benefit from a “black-box” tool for impact measurement. AFD has thus decided to pursue and diversify its investment in methodological approaches and knowledge production in the field of impact evaluation, notably through pilot operations and specialised partnerships. It is also actively engaged in the international debate on this issue.

## 1.2. Impact evaluations are an important part of the evaluation activity

In line with these new orientations, AFD has been reforming its evaluation function since 2006. The reform process is underpinned by a two-pillar system: decentralisation (towards local agencies) of systematic external evaluation of individual financing, and reorientation of the Evaluation Unit towards evaluation quality, strategic evaluation and knowledge production, including impact measurement.

Prior to 2006, the evaluation function had relied primarily on in-house project evaluation, driven and implemented by the Evaluation Unit and placed under the direct authority of the Head of Strategy. About 15% of AFD-financed projects were subject to final evaluation. Project evaluation was presented and discussed in the Evaluation Committee, chaired by General Management. The overall assessment concluded that AFD did not make the most of its evaluation system; the Evaluation Unit was isolated from the rest of the institution;

feedback was weak; and it was difficult to mobilise the Evaluation Committee.

At the beginning of 2006, the Evaluation Unit was thus integrated into the Research Department (the equivalent of Knowledge Department in some institutions), which is part of the Strategy Branch. This original arrangement expresses a clear decision to establish a link between knowledge production and evaluation. Furthermore, it is stipulated in the evaluation principles that part of the evaluation work should include social science research.

This revitalisation of evaluation is grounded in four principles:

1. The sharing of evaluations: evaluation should not be a restricted activity, centralised in a dedicated unit, with the sole purpose of informing top management. Evaluation reports should directly involve all development actors and, in particular, operational departments and local partners.

2. Synergy between evaluation and research: part of the evaluation work is to be research applied to analysing the processes and results of projects and programmes.

3. Focus on capitalisation of experiences and institutional learning: evaluations designed to feed this process are formative rather than summative, are centred on professional practices rather than only policies and strategies, and include ongoing operations.

4. A gradual mixing (and sometimes merging) of external and internal studies: whilst external evaluation should be the common practice, internal evaluation remains necessary not only in order to make the process of institutional learning effective, but also to update know-how within the Evaluation Unit and make the unit professionally attractive.

AFD has recently adopted a number of new evaluation tools: decentralised evaluations, strategic evaluations, thematic capitalisation and rigorous impact evaluations. Decentralised evaluations will be commissioned by geographical departments and local agencies, shared with local

partners and entrusted to external (and preferably local) experts. Strategic evaluations continue to be commissioned and piloted by the Evaluation Unit on the initiative of management and supervisory ministries. The Evaluation Unit is also developing the thematic capitalisation of evaluation, grounded on comparative analysis of clusters of completed or ongoing development operations that are financed by AFD with or without other partners. Rigorous impact evaluations are carried out in collaboration with specialist academic teams and interested local partners.

Since its integration into the Research Department, the Evaluation Unit has received a mandate to develop impact evaluations. The link between knowledge production and evaluation is, in effect, a key factor for facilitating investment in impact evaluation. A serious impediment to developing impact evaluation within evaluation units may well be the frequent formal separation between research and evaluation activities, as well as the common misconception that social science research and evaluation form two disconnected disciplines.

As mentioned earlier, the AFD impact evaluation programme will be developed progressively, which to some extent implies an internal intellectual investment in the impact issue. In broad outline, the programme is pursuing the following objectives: the first goal is strategic and aims at producing sound knowledge about what does and does not work in development policies for Southern partners, for AFD sectoral policymakers and, more generally, for the development community. Secondly, AFD is pursuing a methodological goal to be achieved by reinforcing in-house mastery of various impact measurement tools. The final objective is to foster partnerships through joint knowledge-production efforts with various Southern partners and in active collaboration with scientific expert teams.

Finally, this stepwise experience for AFD will be shared and enriched through active participation in international networks on impact evaluation and by pooling results and methodologies.

## 2. A set of experiences

### 2.1. A preliminary phase

AFD began developing sound impact evaluations as early as 2003 by financing research into the impact of multi-donor agricultural development projects on farmers' income in Guinea (Delarue, 2007). This impact evaluation was conducted by Agroparistech<sup>1</sup> with a view to proposing a methodology suitable for quantifying impacts using a qualitative approach. It was carried out entirely *ex post* and relied on a structured survey using informant recall to collect data about the pre-intervention period, and on in-depth interviews with one hundred farmers. In order to quantify the net change in the farmers' income produced by one of the projects, the researcher identified a set of farmers that were directly or indirectly involved in the project as a credible comparison group. Two types of project were evaluated: inland valley development for irrigated rice cultivation and a public agro-industry producing rubber and palm oil.

An in-depth study of a limited number of production units enabled the evaluator to identify a typology of production systems extant before the project. In order to set up a counterfactual, a judgement sample was then obtained by choosing production units that belonged to the same initial type of production system and that had developed with or without the project.

A detailed understanding of the endogenous and exogenous factors influencing the evolution and potential trajectories of production systems helped the evaluator to rigorously identify those individuals with comparable trajectories, whether or not

they were exposed to the project. The evaluator's direct involvement in data collection was essential—hence the importance of a small sample. It would not have been possible to gather such reliable data on yields, modifications to production structures over time and producers' strategies from a large survey sample in a rural context.

This understanding of the projects and trajectories of the farmers (whether or not they were exposed to the project) made it possible to build a quantitative model based on Gittinger's method of economic analysis of development projects (Gittinger, 1982). As the initial diversity of production units was clearly identified before sampling, the model was constructed for each type of farming system extant before the project. Understanding the potential development paths for each farming system, whether or not exposed to the project, allowed the differential impact of the project on the farmers' incomes to be estimated.

The objective differences between each production unit studied might seem to give some leeway to the researcher's subjectivity for the construction of the typology and sample. However, the rationale behind the production system concept made it possible to transcend the risk of arbitrariness. In fact, the methodological jump from a small number of interviews to a model is bridged by demonstrating that a finite number of production system types exist in reality.

The primary interest of this new method is the fact that it provided the opportunity of building a credible impact assessment entirely *ex post*. Secondly, it gave an estimate of the impact on different types of farming systems, thus explicating the uneven distribution of the projects' benefits. Thirdly, it permitted a

<sup>1</sup> AGROPARISTECH is a member of the Paris Institute of Technology which is a consortium of 10 of the foremost French Graduate Institutes in Science and Engineering. It is a leader in Life Sciences and Engineering.



nuanced understanding of why the desired impacts materialised or not.

The results from this first impact assessment became available after four years of field work and data treatment. They were presented to the Guinean authorities and the local representatives of the main donors in the rural sector. In the field, the results were delivered to the local communities interviewed and to the farmers' syndicates. The Minister of Agriculture declared that he would try to foster further impact evaluations on agricultural development projects. Unfortunately, in the absence of an institutionalised forum for discussions between the different stakeholders, there is little hope that the conclusions of this research will change national policy on these types of projects.

## 2.2. The second impact assessment financed by AFD

The following evaluation concerned AdéFI, a microfinance institution (MFI) serving micro-entrepreneurs in Antananarivo (Madagascar). It was launched following a request from both the AdéFI management and AFD to produce valid project data and analyse these using a scientifically robust method. The impact evaluation was conducted between 2003 and 2005 by researchers from IRD-DIAL, a French research centre (Gubert et Roubaud, 2005).

Initially, the methodology involved comparing the situation of a representative sample of AdéFI micro-enterprise clients with a comparison group, constructed through a standard matching technique (propensity score matching). This first quantitative impact assessment was based on a "post-test project and comparison groups" evaluation design (Bamberger et al., 2006). It relied on 255 interviews conducted in 2001 and was, at the time, complemented by a qualitative analysis based on open interviews with a limited number of AdéFI clients.

This analysis proved quite encouraging as far as the project's target group was concerned (in line with the project's theory), and in terms of impacts (on the client enterprises' turnover or production). This first design, however, was not sufficiently robust and a second phase was programmed so as to incorporate a double-difference technique. This second phase, completed by two successive surveys on the same panel in 2003 and 2004, enabled information to be gathered on the dynamics involved. It also enhanced the quality of results by allowing for the inclusion of new variables in the matching process, thus rendering clients and non-clients even more similar than in phase one.

Unfortunately, it was not always possible to keep track of the enterprises in the 2001 treatment and comparison groups: in 2003, the attrition rate for the two groups was respectively 22% and 23%. In 2004, only 55% of the enterprises from the original panel were still active. This low survival rate evidences the great fragility of small enterprises and, against expectations, AdéFI's clients were affected to a greater extent. Whereas 255 enterprises were interviewed in 2001, only 107 interviews could be used for the panel analysis in 2004.

Other methodological aspects were improved during the second phase. The observables selected for the propensity score regression (probit) for both phases included the micro-entrepreneur's gender, age, educational level, type of learning, the enterprise's economic sector, the type of premises in which the activity is exercised, the creation date of the enterprise, the initial workforce, the initial value of capital stock, etc. In 2001, this information had been requested with respect to the year the enterprise was created, which was not identical for all of them. In 2003 and 2004, this information was requested for 1997 in order to control for the differences in prevailing characteristics between clients and non-clients at the moment when AdéFI started.

Several variables of interest relating to the economic performance of the micro-enterprises were studied: turnover, production, added value, workforce, capital and, finally, producti-

vity of labour and capital. The impact of micro-credit on these variables appears to be positive and statistically significant in 2001 and 2004. Yet, overall, the impact measured in 2004 seems to be smaller than the impact that was first assessed using the 2001 data. This difference can be partly explained by the methodology used: the 2004 data-matching was more rigorous and the 2001 results probably contain a bias.

In fact, because the matching in 2001 was based on the characteristics of the enterprises in the year they were created, the propensity score was not based on output variables (for instance, turnover, production, etc.). In 2004, on the other hand, their 1997 turnover was included in the variables used in the model to predict participation, which helped to obtain a better match. A simple comparison between the use of the two sets of variables for calculating the propensity score on the basis of 2004 data showed a significant difference in the impacts measured, even though the impact remained positive for all variables. This test demonstrated both how very important and how very difficult it is to build an adequate comparison group, particularly when no baseline study has been carried out.

Finally, the use of the double-difference technique between 2001 and 2004 gave very different results from the aforementioned single-period measures. With the latter, the project showed a positive impact on productivity and the different outputs. In contrast, when the double-difference technique was applied, none of the measured impacts were significant. This means that the path of economic performance for clients and non-clients is on average identical, and that the project has not succeeded in activating a growth dynamic for its clients.

This impact evaluation showed how difficult it is to collect panel data on the clients of a micro-finance project. In the present case, this was due to the high attrition rate, which is closely tied to the vulnerability of the micro-enterprises and to their propensity to change location, thus obliging interviewers to keep track of them, often in vain. This evaluation also evidenced the highly sensitive nature of the relationship between impact measures and the quality of matching.

This study is one of the rare impact evaluations to assess the impacts of a micro-finance institution on micro-entrepreneurs. Completed in 2005, it is also AFD's first experience in conducting impact evaluations. The rigour and transparency with which the research team carried out the scientific work has made a positive contribution to AFD's institutional learning in the field of impact evaluations. It has also encouraged AFD to foster new impact studies that address the methodological limits of this first experience, and led to more substantial budget allocation for impact evaluation. In particular, subsequent impact evaluations have been programmed far in advance of project start-up, with comprehensive baseline surveys to ensure collection of the necessary information about the initial situation.

### 2.3. A randomised controlled trial of micro-finance in Morocco

The first experimental impact evaluation financed by AFD involves Al Amana, a micro-finance institution that is active in rural areas in Morocco.

There still exists an "evaluation gap" concerning micro-finance programmes: *"MFI field operations have far surpassed the research capacity to analyze them, so excitement about the use of microfinance for poverty alleviation is not backed up with sound facts derived from rigorous research. Given the current state of knowledge, it is difficult to allocate confidently public resources to microfinance development."* (Zeller and Meyer, 2003). Moreover, even though micro-finance has been the subject of a good many impact evaluations over the last ten years, this is one of the first to use a randomised controlled trial. It is particularly interesting to measure micro-credit impact in rural areas, as reaching one of the poorest categories of population constitutes a real challenge for many micro-finance institutions.

Al Amana, which was created in 1995, is the largest micro-finance institution in Morocco and serves some 250,000 clients. Until 2006, the institution's clientele was located prima-

rily in urban or peri-urban areas (83% of clients), although its current strategy targets wide-spread rural areas. After opening approximately 100 branches in easily accessible hinterlands in 2004 and 2005, Al Amana decided to expand into the isolated rural regions. Keen to obtain rigorous measures of the impact of micro-credit distribution in this challenging new context, Al Amana management decided to ask AFD for financial support to conduct a study. The institution had already identified the Poverty Action Lab (MIT) as the research team that would be in charge of the evaluation, in partnership with the newly created Paris School of Economics.

The objective of the research programme is to analyse the economic and social impacts of micro-credit in isolated rural regions in Morocco, using an experimental method (Paris School of Economics, 2006). The randomisation of the treatment assignment – with one group being exposed to micro-credit from the outset and another group at a later date – will give clear, transparent and rigorous estimates of the impacts. The roll-out of Al Amana's activities provides an ideal context for using this type of method.

The evaluation concerns 80 of the 160 branches that Al Amana has planned to open between 2006 and 2008. The principle of the study is to identify two small zones within the area covered by a branch, one zone being served quickly and the other one year later, as initially decided. In this scenario, three surveys were to be conducted: a baseline survey, an intermediary survey after one year and a final survey after two years. The final survey was to allow measurement of the effects of two years of credit distribution compared to one year in the control group, thus providing a differential analysis of short-term and medium-term effects of the treatment on the populations.

The *modus operandi* for setting up the two groups was defined on the basis of a feasibility test first carried out on nine sites spread around Morocco. It was not possible to select the villages by simply drawing from a list, as there is a great diversity of rural settings, notably in terms of the type of land

tenure, crops, landscape and climate. Moreover, it was important that the douars (Moroccan villages) in the control group be distant enough from both the place where the micro-credit branch was established and from all other source of credit in order to minimise the risk of contamination. The villages in the treatment group had therefore to be chosen within similar contexts. The feasibility survey helped to define a matching method for choosing a pair of villages (treatment and control) with the same characteristics. The variable used for the selection included accessibility, population, main crops cultivated, etc. The draw was eventually completed with each branch having a pair of similar villages, one village being randomly assigned to the treatment group and the other to the control group.

Feasibility played yet another crucial role in defining the procedures for the experiment, notably in constructing a model to predict the villagers' propensity to take up credit. In order to limit the number of interviews needed to reach statistical power, it was vital to select as many future borrowers as possible from among the households interviewed during the baseline survey. During the feasibility study, 2,000 households in the nine pairs of villages were interviewed and the distribution of credit over the subsequent six months was tracked. It was this phase of observing the credit take-up that enabled the predicting model to be proposed.

The model made it possible to forecast which 25 households in the remaining villages had the highest propensity to take up credit with Al Amana, on the basis of a short questionnaire (10 questions) applied to 100 families.

The data collected at the first nine sites showed that the evaluation process was progressing successfully. The randomisation was working satisfactorily and the original differences between the households in the treatment and the control groups were not significant. Moreover, the collaboration between Al Amana and the research team was exemplary. Nevertheless, several technical problems arose and led to a change in methodology.

The surveys conducted during the first year revealed that take-up was lower than was theoretically expected in the enclosed regions where Al Amana had no previous experience. For the first branches involved in the feasibility study, the borrowing rate was 21% after 14 months (36% amongst the households of the treatment group, thanks to the propensity model). In the 23 branches subsequently included in the study (representing 39 douars in the treatment group), the average take-up rate was only 7% after 5 months, with enormous variations between villages (from 0 to 55%).

The lack of borrowers in the treatment group interviewed put the final possibility of measuring a limited impact with statistical significance into jeopardy. With a borrowing rate of 20%, it would be impossible to detect a change in consumption smaller than 21%. In order to address this problem for the needs of the evaluation, various steps were taken to increase the awareness of villagers in the treatment douars: the number of information meetings was increased; the quota of credit reserved for women was opened to any borrower; and incentives were given to Al Amana staff to serve these remote villages.

As it happened, the protocol had to be revised more profoundly. The setbacks encountered indicated that the one-year exclusion period would not be long enough to obtain a significant difference between the treatment and control groups. It was thus decided to extend the exclusion period of the control group to two years. This decision is not without consequence. For the local Al Amana staff, it means explaining to the population of the control village that the access credit has been delayed. It also means that Al Amana is serving fewer clients, which in turn leads to a shortfall for the institution. Since the mid-term survey had been cancelled, its budget was reallocated to include 20 more villages in the survey so as to have a greater chance of reaching final statistical significance.

These substantial adjustments have been made possible thanks to the remarkable partnership that has built up between Al Amana, the Paris School of Economics and AFD in order to overcome the difficulties encountered and ensure completion

of the evaluation. The meetings between all the parties are regular and help to prevent any misunderstanding. The research protocol is highly transparent for all the stakeholders and results will be available in 2010.

## 2.4. A randomised controlled trial of micro health insurance in Cambodia

Health insurance is one of the most salient policy issues facing the developing world today, and France has recently pledged to step up its investments in social protection for developing countries. AFD is a relative newcomer to health programme financing and sustains only two micro health insurance programmes: one in Cambodia and one in Laos.

In order to reinforce its knowledge in this area before scaling up investments, AFD decided to launch an impact evaluation of the SKY Health Insurance Programme which it finances in Cambodia. Launched in 1998 by the *Groupe de recherche et d'échanges technologiques* (Research and Technological Exchange Group: GRET), SKY offers households free and unlimited primary and emergency care at health centres, as well as a number of other health services, for a fixed monthly premium. One of SKY's primary goals is to enable families to cover health costs without them being pushed into poverty.

In 2005, AFD signed a memorandum of understanding for the execution of project evaluations with the Scientific Evaluation and Global Action (SEGA) research centre run by the University of California, Berkeley, and the University of California, San Francisco. Further to a methodological proposal written by SEGA for impact evaluation of the SKY micro-health insurance project, an identification mission took place at the end of 2006. This aimed to define more precisely the possible methodology and scope of the evaluation, and to start fostering the buy-in of future findings by policymakers in Cambodia.

At the core of the impact evaluation of the SKY project is a randomised controlled trial (Levine et al, 2007). It will be implemented in parallel to SKY's roll-out in the Takeo province, currently scheduled to begin around June 2008. In the preferred study design, the central methodological tool is the use of randomisation of insurance premium levels in order to vary the likelihood of insurance take-up among households within a village and isolate the impact of health insurance on the outcomes of interest.

Following an initial village meeting, when coupons for premium reductions will be randomly distributed, the baseline survey will be administered to a random subset of households, stratified by coupon value. Using the baseline survey data and SKY's records of which households opted to take up insurance, it will be possible to answer the questions about selection into the insurance programme. For example, it will reveal which household characteristics predict take-up of health insurance. Furthermore, since the premium is randomly assigned, it will be possible to assess how premiums affect the baseline characteristics of insured versus uninsured households.

Twelve months after each village meeting, follow-up surveys of all the households originally interviewed will be carried out. The follow-up and the baseline data will provide information on how SKY affects health-seeking behaviour and healthcare utilisation, as well as on how health insurance influences economic outcomes, such as changes in out-of-pocket expenditures. One year later, a second follow-up will repeat most of the same topics, again emphasising changes in health outcomes and expenditures.

Since longer-term effects of insurance are also very interesting, high drop-out rates among large-coupon winners is of concern. If pilot tests show that most people who win a coupon in the first period renew their insurance for an additional 6 months, the above design will suffice for effects over at least the first 12 months.

However, as in any project, the evaluation might not go entirely according to plan. The main risk lies in the number of households that do not continue their membership with SKY after their initial six-month period. Currently, the drop-out rate for SKY after the initial six-month period is approximately 17% (based on past SKY records). If drop-out is substantially higher than this for purchasers who received high-value coupons, there could soon be little difference between the insured status of the initial Treatment (high coupon) and Control (low coupon) groups. It may then be necessary to administer a non-experimental "matching" method to gauge the impacts of SKY.

In addition to household surveys, the research team will also administer a qualitative evaluation of the SKY programme. This analysis will examine the impacts that SKY has on the health system, including public health facility revenue, changes in supply of drugs and medical equipment, and changes in health workers income and work patterns.

Finally, randomisation will allow the researchers to credibly estimate the causal effects of health insurance, as distinct from all other characteristics that vary across insured and uninsured households. A pre-intervention baseline survey of approximately 3,000 households, involving over 15,000 individuals and follow-up surveys of the same households, will be conducted over the 4-year experimental period. The survey will cover the multiple areas that the programme aims to influence: health status, health-seeking behaviour, asset vulnerability, investment and saving decisions, and risk management. Drawing upon the randomised research design, it will be possible to compare the changes in outcomes over time across insured and uninsured households in order to estimate the causal effect of health insurance.

A key feature of this impact evaluation is the series of partnerships both within Cambodia and at international level so as to ensure that the evaluation adequately matches the needs of the programme, the funders such as AFD, and other stakeholders with an interest in health care delivery to the poor.

Throughout the drafting of the methodological proposal, SEGA worked closely with staff from AFD and GRET, as well as Cambodia-based research partners from Domrei Research and Consulting. GRET's input was particularly crucial for all aspects of the proposal, including determining the feasibility and relevance of research designs. These relationships are further developing as the research design is being structured and implemented and as the survey instruments are being installed.

In addition, SEGA has been awarded a USAID-funded BASIS grant, which will allow it to develop capacity-building activities for local researchers and practitioners in Cambodia. These activities involve partnering two researchers from the Royal University of Phnom Penh (RUPP). Throughout the evaluation period, SEGA will offer RUPP students training in programme evaluation design and methodology, with the aim of enabling future impact evaluations to be run locally. Moreover,

part of the budget allocated by AFD includes presentations in Paris to disseminate SEGA's methodologies (including research design and econometric techniques) and findings.

As was confirmed during the first mission to Cambodia, the evaluation of the SKY programme directly supports the goals of the Cambodian Ministry of Health. In particular, knowledge about SKY's effectiveness will help the Ministry to structure its reform of the Cambodian health care system, which has been ongoing since 1999 (ILO, 2005). During this mission, the evaluation team and AFD met with Ministry of Health officials to collect information on questions that they consider particularly important to address in the SKY evaluation. Throughout the evaluation, the input of Ministry of Health officials will be sought regarding the study design and they will be regularly informed on results.

The results will be available in 2010.



### 3. Intermediary lessons

#### 3.1. Institutional learning

All of these experiences have enabled AFD to progressively learn and change its approach to impact evaluations. The main outcome has been the adoption of methods that require baseline surveys to be carried out. This means it has been necessary to convince the various stakeholders of this necessity and to identify the academic partners far in advance of the project launch. Preparing such research is obviously time-consuming, as it involves getting policymakers, researchers, data collectors, development operators and donors to meet and agree on the principles and details of the exercise. In the Moroccan and Cambodian cases, it meant dedicating one and a half year to the preparation phase.

Progress has also been gradually achieved in terms of the implication of local partners. It is very clear that impact evaluations constitute a very demanding exercise that requires the full involvement of the project team being assessed. The project team's contribution is essential for contextualising the questionnaires and adjusting the sample size according to the expected take-up. The project team's implication is especially crucial during the impact evaluation's implementation phase in order for the design to be correctly applied, particularly when it comes to preventing contamination of the control group. It also often means that the project and its various stakeholders must be ready to change the intervention protocol to allow for randomisation or for constructing a good comparison group.

Yet, the strategic objective of implementing impact evaluations as a means of contributing to policy-making implies more than involving the project team. Rigorous findings may not

make any difference if they fail to be taken into account at the level of country policymakers. Particular efforts were made in the Guinean and Cambodian evaluations to promote a political buy-in from start to finish.

Based on the current experiences, impact evaluations appear as much a research product as an evaluation product. As far as their expected feedback is concerned, they are close to evaluation but the nature of investigation is clearly one of research. They have created new opportunities for AFD to collaborate with high-quality academic partners in analysing its operations. A high level of in-house involvement in each evaluation, from project managers and the Evaluation Unit alike, is developing an Evaluation culture that is favourable to launching more evaluations of a "scientific" quality in the future.

#### 3.2. Impact evaluation for a bilateral donor

AFD's track record in impact evaluation might appear minimal: the exercises that will be in progress in 2008 (Cambodia and Morocco) represent only two of about five hundred ongoing AFD-financed operations. However, these two projects will consume 25% of the Evaluation Unit's budget for out-sourced evaluations and between 10% and 15% of its human resources. This disproportion deserves attention.

For a bilateral donor, the first question concerning impact evaluation (IE) is to wonder if it is worth engaging in such efforts. The second question is whether to internalise or out-source the piloting of this kind of work. If the internal option is preferred, then the third question arises of "where to locate the

IE activity: in the evaluation unit, knowledge department or policy department?"

As explained above, AFD has endeavoured to address these questions by launching several impact evaluations piloted by its Evaluation Unit. The last question concerning IE location was not an issue for the AFD given that the Evaluation function is now incorporated into the Research Department.

With the hindsight of AFD's brief experience, internal involvement in IE piloting appears to be an effective way of fuelling the debate on impact measurement within the institution. Moreover, developing results-based management could hardly be envisaged without experiencing the implications of rigorous impact assessment.

Direct implication in the management of IE is crucial to fully understanding the dos and don'ts and the objective difficulties and challenges of rigorous impact evaluations, as evidenced by the aforementioned examples. Internalising IE or outsourcing to a pool dedicated to this activity could in fact be a false alternative. Appropriation of IE is a prerequisite for fruitful participation in international networks, with sharing not only of results but also of processes and methods.

### 3.3. Perspectives

Our analysis of the Randomised Controlled Trials (RCTs) launched so far shows that, although these trials are very interesting, they present some limitations when applied to the type of projects that AFD finances. A recurrent difficulty when a quantitative methodology with a baseline study is applied to AFD projects is the self-selection of beneficiaries, as well as the slow take-up rate when new specific services (e.g. micro-credit or micro-insurance) are proposed to populations for the first time. Moreover, there is a contradiction between maintaining a contamination-free control group and measuring the impacts of a new project – which may take time to materialise – on individuals and on a population as a whole.

The results of the recently launched RCTs are not expected until few years from now. Each of the trials required an eighteen-month identification phase prior to start-up. Due to their cost, the demanding preparation and implementation work involved, the aforementioned objective risks in evaluation design and their unknown impact on policymaking, AFD has decided to wait for the results before engaging in another exercise of this type.

One of AFD's concerns for the future would also be to identify methods that provide answers to a maximum number of relevant questions at a policy level. To this end, a mix of methods have been or will be used in the Guinean and the Cambodian evaluations: information relating to processes and impacts on key stakeholders other than the direct clients of the project will be collected and analysed. Nonetheless, it remains difficult to find researchers who are spontaneously open to using a necessary combination of methods for impact evaluation. This means dedicating a great deal of effort to fine-tuning the main method used. Both the project and AFD have a responsibility to ensure that providing answers to the maximum number of relevant questions prevails over method-oriented research.

Indeed, AFD's Evaluation Unit has now taken up the challenge of developing innovative approaches to rigorous impact evaluation methodology, by using counterfactuals and by quantifying the net change in outcomes without necessarily using experimental or quasi-experimental techniques. A variety of rigorous impact evaluation methods is essential in order to bring answers to the array of questions raised by development interventions, whilst also addressing the question of attribution. As evidenced by the Guinean impact evaluation, the AFD Evaluation Unit believes that qualitative research can rigorously address the challenges of impact evaluation.

To face the growing in-house demands for impact assessments, AFD plans to adopt a pragmatic approach that not only fosters more quality evaluations in general, but also integrates the attribution question where possible. Few project managers



are ready to finance a baseline study and specific data collection, and even fewer are currently willing to engage in a long-term process and allocate a sizeable budget to a research programme on the impact of their operations. Clearly, more information is required about the real effects of impact evaluations before these can be developed on a wider scale.

## References

AFD (2007), *2007-2011 Strategic Plan of the Agence Française de Développement*, Paris.

Bamberger, M., Rugh, J., Mabry, L. (2006), *Real World Evaluation : working under budget, time, data and political constraints*, Sage publications.

CGD (2006), *When will we ever learn? Improving lives through impact evaluations*, Report of the Evaluation Gap Working Group.

Delarue, J. (2007), *Mise au point d'une méthode d'évaluation systématique de l'impact des projets de développement agricole sur le revenu des producteurs, Etude de cas en région kpélè (République de Guinée)*, PhD thesis, AGROPARISTECH, Paris.

Gittinger, J.P. (1982), *Economic Analysis of Agricultural Projects* (EDI Series in Economic Development), John Hopkins University Press, Baltimore.

Gubert, F. and F. Roubaud (2005), *Analyser l'impact d'un projet de microfinance : l'exemple d'ADéFI à Madagascar*, Notes et Documents, no. 19, AFD, Paris.

ILO (2005), *Informal Economy Poverty, Employment in Cambodia, Mongolia, Thailand: The First Eighteen Months – A Progress Review*, ILO, Geneva. [www.ilo.org/public/english/bureau/integration/download/publicat/4\\_3\\_370\\_informal\\_economy\\_final\\_report\\_-\\_with\\_cover\\_sheet.pdf](http://www.ilo.org/public/english/bureau/integration/download/publicat/4_3_370_informal_economy_final_report_-_with_cover_sheet.pdf)

Levine, D. I., R. Polimeni and R. Arunachalam (2007), *A randomized controlled trial of micro-health insurance in Cambodia, A preliminary proposal to AFD, SEGA, UC Berkeley, Berkeley*.

Paris School of Economics (2006), *Evaluation de l'impact d'un programme de microcrédit en milieu rural*, Research project, Paris School of Economics, Paris.

Paris School of Economics, (2007), *Evaluation de l'impact du microcrédit en milieu rural*, Note written for the Steering Committee 1st August 2007, Paris School of Economics, Paris.

Zeller M. and R.L. Meyer (dir.pub.) (2003), *The triangle of microfinance: financial sustainability, outreach and impact*, John Hopkins University Press, Baltimore.