

# Evaluation in International Development Aid Agencies: the Case of USAID

강 경 재 KOICA ODA연구실 상임연구원

## 목차

- I . Introduction
- II . Evaluation in USAID: In Theory
- III . Conclusion: The Lessons Learned from USAID  
Evaluation Experience

## Abstract

The United States Agency for International Development (USAID) recently announced its new evaluation policy. The policy is the end results of an unending process for organizational renovation. Evaluation at USAID has degraded during the course of years, both in quantity and quality. This paper examines the policies and practices of evaluation in USAID, and presents examples of USAID evaluation. The results of reviewing cases and existing literature suggests that effective evaluation requires: a) an agreed consensus of the agency on the definition and modality of evaluation, b) cultural and structural support of the agency for evaluation, and c) clearly and consistently written guidelines should be provided.

“Program evaluation ... is valuable only when its findings are applied.”

—USAID Evaluation Handbook, 1970

## I . Introduction

On October 28, 2007, an external evaluation team arrived at Kabul Airport. Their mission was to evaluate the United States Agency for International Development's (USAID's) Rebuilding Agricultural Markets Programs that ended in 2006 over the following few weeks using a base method that included reviewing written information, data, and visiting sites in war-affected Afghanistan. Their tasks seemed daunting. Even though the projects had detailed information on its activities and performance, it was an extremely complicated program. There were many external factors influencing the program outcomes and impacts. The sites the team could visit were very limited due to security reasons. Despite all the difficulties, the evaluation was carried out and the results were reported.<sup>1)</sup>

This was how a typical evaluation was carried out. An evaluation team, either USAID internal evaluators or external experts, flew to the site, spent a few weeks for data collection and dissemination, and finally delivered a report. On the surface, this procedure remains largely the same. However, the emphasis on and organizational support for evaluation have changed significantly over the last decade. In 2007, when this impact evaluation was carried out, evaluation was reemerging as a priority in USAID after it had dropped out from the top 10 list of years with the highest number of USAID workers since the mid-1990s. In 2008, a central evaluation unit was established. Finally, on January 19, 2011, USAID announced a new evaluation policy. The policy is considered to be a big step forward. According to Levine, Deputy Assistant Administrator of the Bureau of Policy, Planning and Learning at USAID, the new policy “sets a new standard for evaluation practice.”<sup>2)</sup> The key points of the new evaluation policy include distinguishing impact evaluation from performance evaluation, designating sufficient funding for evaluation, and incorporating evaluation into the design of all programs and policies. Also, performance evaluation has become mandatory for all operations<sup>3)</sup> in USAID.

---

1) USAID (2007), “Impact Evaluation: Rebuilding Agricultural Markets Program,” Washington, D.C., USA

2) Ruth Levine (2011), “USAID's Evaluation Policy: Setting the Standard,” <http://blog.usaid.gov/2011/01/usaid-evaluation-policy-setting-the-standard>, accessed May 15 2011

USAID was among international development agencies that adopted a results-based management (RBM) approach, and the interpretation of RBM affected the evaluation practices.<sup>3)</sup> In the past, the emphasis on the performance monitoring turned evaluation into a non-essential component in management. However, currently USAID considers evaluation as a key tool in “management for development results,” the latest innovation in enhancing effectiveness of the development aid. In its new evaluation policy, USAID declared that “USAID bases policy and investment decisions on the best available empirical evidence and uses the opportunities afforded by project implementation to generate new knowledge for the wider community.” Evaluation is also key feature in addressing accountability.” Since all international development agencies’ budgets come from the taxpayers, the agencies are obligated to explain their financial and performance management to parliaments, and to taxpayers.<sup>4)</sup> This forces development agencies to measure and prove their performance and impact of actions and programs.<sup>5)</sup> Therefore, all international development agencies, including USAID and the Korea International Cooperation Agency (KOICA), are on a quest for better evaluation.

In January 2011, while USAID announced its new policy on evaluation, on the other side of globe, KOICA was also building the agency’s capacity in evaluation. KOICA’s past experience in performance evaluation was examined and considered only partially satisfactory. The accountability issue was (and still is) building up; the official development assistance (ODA) budget of Korea is going to be significantly increased as a response to pressure from the Organisation for Economic Co-operation and Development’s Development Assistance Committee (OECD DAC).

This paper is the result of a seminar held at KOICA on March 30, 2011. The seminar series is a part of efforts to improve evaluation performance and practice at KOICA. The aim of this paper is to provide some insights on evaluation practices by reviewing USAID’s evaluation policy and practices.

---

3) Ryoh Sasaki (2006), “A Review of the History and the Current Practice of Aid Evaluation,” Journal of Multidisciplinary Evaluation, Number 5, pp. 55-88

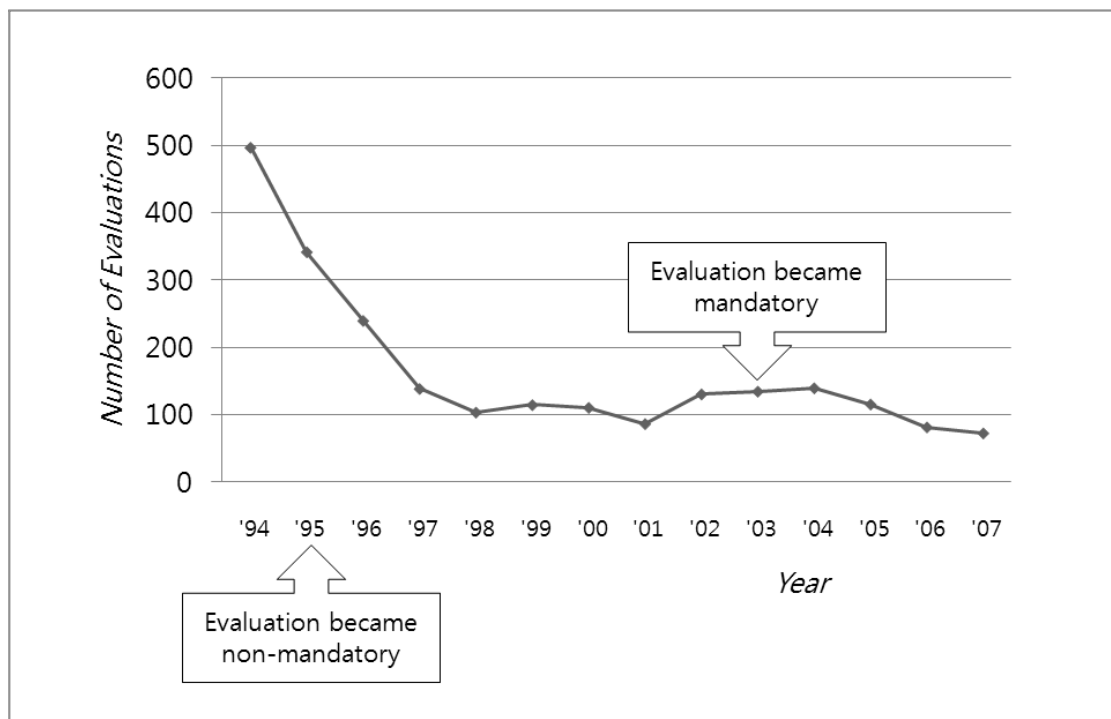
4) Howard White (2005) , “Challenges in Evaluating Development Effectiveness,” Institution of Development Studies, Brighton, England

5) Faith Corneille (2001), “Entangled Accountability in Development Agencies: The Case of USAID”

## Evaluation at USAID: The Past

To move forward, one must know the past. USAID invented and nurtured the practice of evaluation. From the very beginning, USAID actively invested in evaluation. As a result, USAID invented various techniques and concepts to enhance evaluation results. For instance, USAID developed the concept of impact evaluation in the 1980s. However, it also often suffered from evaluation “fatigue.” Often, the lessons learned were not sufficiently shared among the members of USAID. As a result, evaluation was pushed away from the agency’s priority in 1994; evaluation was extra activity which was used on an as-needed basis, rather than a mandatory basis. The shift was caused by USAID’s emphasis on performance monitoring systems. USAID developed and maintained a strong performance monitoring framework, which provided an effective tool for quality data collection on the agency’s performance. However, such emphasis on performance monitoring undermined the role of evaluation in management and decision making.<sup>6)</sup>

〈Figure 1〉 Number of Evaluations from 1995 to 2007



6) OECD DAC (2006), “The United States DAC Peer Review”

As Figure 1 shows, the number of evaluations between 1995 and 2007 dropped significantly. This drop occurred due to an actual decrease in the number of evaluations and a lack of reporting/sharing of evaluation reports. Many evaluations were carried out by implementation partners, which was common practice in USAID. However, unlike before 1994, many partners and implementing staff either did not finish their evaluations or did not submit them to the Washington office. The logic for USAID to change the evaluation from a mandatory requirement to that which is performed on an as-needed basis was: a) the agency chose not to fund individual projects but to support strategic plans and programs; and b) it had adopted results-based management and put emphasis on performance monitoring, which is considered to be a sufficient tool for obtaining information.<sup>7)</sup> The drop in the number of evaluations did not stop until 2007. In 2007, the number of evaluations registered in the Development Experience Clearinghouse<sup>ii)</sup> was 73, less than one-fifth of what it used to be.<sup>8)</sup>

Such a drastic drop attracted USAID's concern and triggered several large-scale studies and an evaluation on the evaluation process itself. Also, in 2003, all programs again became subject to obligatory evaluations, and in 2005, USAID Administrator Andrew Natsios launched an agency-wide Initiative to Revitalize Evaluation. Natsios said, "Evaluation is at the heart of three of the nine principles of development that guide Agency operations: the principle of accountability, the principle of assessment, and the principle of results."<sup>9)</sup> However, this ambitious initiative did not last and the number of evaluations kept decreasing until 2007. The call for the revival of evaluation was getting louder. In 2008, as part of efforts to renew evaluations, a central evaluation unit was created and evaluation is now considered to be a priority in the new evaluation policy. Meanwhile, various studies and evaluations pointed out problems in the evaluation of U.S. foreign assistance. The reviews on USAID performance repeatedly recognized the underutilization of evaluation as a tool for "learning" and "accountability."<sup>10), 11), 12)</sup> Changes were made in the institution. Many rules

7) Cynthia Clapp-Wincek and Richard Blue (2001), "Evaluation of Recent USAID Evaluation Experience," USAID PPC/CDIE, document PN-ACG-632

8) USAID (2009), "Trends in International Development Evaluation Theory, Policy and Practices," Washington, D.C., USA

9) USAID (2005), "Note from Natsios" Frontline, March, p. 3

10) Faith Corneille (2001), "Entangled Accountability in Development Agencies: The Case of USAID"

11) Cynthia Clapp-Wincek and Richard Blue (2001), "Evaluation of Recent USAID Evaluation Experience," USAID PPC/CDIE, document PN-ACG-632

12) Janice M. Weber (2004), "An Evaluation of USAID's Evaluation Function: Recommendations for Reinvigorating the Evaluation Culture within the Agency," USAID

and directives were adjusted and a central evaluation unit was established. Finally, USAID announced a new policy in January 2011, which reflected the recommendations and findings suggested by previous works. Now, evaluation is being considered in the implementation and management of foreign assistance agencies once again.

## II. Evaluation in USAID: In Theory

USAID evaluation is guided and regulated by both internal and external regulations and directives. The most fundamental one is the Government Performance and Results Act (GPRA) of 1993. GPRA requires U.S. Government agencies to have: a) five-year strategic plans, each of which contains a mission statement for the agency and long-term results-oriented goals; b) annual performance plans including goals and the methods to achieve the goals and to verify achievement during the fiscal year, and c) annual performance reports containing performance reviews (indicating either success or failure). Even before the GPRA, USAID acknowledged the crucial role of evaluation in efficient management and had various recommendations and internal directives in place to aid the evaluation process.

In this section, “Evaluation Guidelines for Foreign Assistance”, “Evaluation Policy” and “ADS<sup>iii)</sup> Chapter 203” three major documents on USAID evaluation guidelines are described, in order to help understand the principles and procedures of USAID evaluation.

### USAID Documents on Evaluation Policies and Guidelines

Guidelines for USAID are provided in several documents: “Evaluation Guidelines for Foreign Assistance”, “Evaluation Policy”, “ADS Chapter 203” and various documents such as “TIPS.” The first three are on the policies and regulations regarding evaluation, while the last provides technical details which are applicable for program planning, implementation and evaluations. For the interest of this research, only the first three are examined.

“Evaluation Guidelines for Foreign Assistance” of the State Department’s Office of the Director of Foreign Assistance (DFA) and USAID provide general guidelines for the evaluation

of all foreign assistance agencies. Established in 2006, the DFA oversees all U.S. foreign assistance programs.<sup>13)</sup> The content of the Evaluation Guidelines for Foreign Assistance is not newly invented; since the 1970s, USAID has accumulated tremendous amounts of guidelines, reports and literature on evaluation. Utilizing this rich pool of resources, the DFA based the “Evaluation Guidelines for Foreign Assistance” on the previous efforts of USAID and set the guidelines to be followed by other U.S. government foreign assistance agencies.

“**ADS Chapter 203: Assessing and Learning**” contains specific details of evaluation, clarifying mandatory and non-mandatory components of program evaluation.<sup>14)</sup> The distinction between mandatory and non-mandatory guidance is made in order to guide users clearly on what “must be done” or “should be done” in order to achieve an ideal quality of evaluation. Mandatory guidelines are based on U.S. law, regulations, USAID policies, directives and contracts. Evaluation itself is now considered to be a mandatory component. The detailed information contained in ADS Chapter 203 is summarized in Table 1.

〈Table 1〉 Summary of ADS Chapter 203

ADS	Details	Mandatory
203.3.2 Performance Management	Processes, principles, and budgeting for performance management	None
203.3.3 Performance Management Plans (PMPs)	Content and format of performance management plans	Complete a PMP for each Assistance Objective (AO)
203.3.4 Selecting Performance Indicators for PMPs	Types, characteristics of performance indicators, gender consideration, reporting requirements, setting baselines and targets, updating PMPs and changing performance	Inclusion of performance indicators in PMPs, gender-sensitive indicators for gender inequality interventions and the expectation that interventions affect women and men differently
203.3.5 Data Quality	Quality standards, purpose and process of data quality assessment	Data quality assessment must have been conducted within the three years before submission for data reported externally or for GPRA reporting purposes

13) Office of the Director of U.S. Foreign Assistance (2009), “Evaluation Guidelines for Foreign Assistance”

14) USAID (2010), “ADS Chapter 203: Assessing and Learning”

ADS	Details	Mandatory
203.3.6 Evaluation	Planning an evaluation, a statement of work, methodologies, participation, documenting, responding and sharing an evaluation	At least one performance evaluation/AO each year through the submission of a finalized (if not finalized, the last draft of the) evaluation to the Development Experience Clearinghouse (DEC)
203.3.7 Portfolio Reviews	Issues to be addressed by, illustrative questions for, and documentation of portfolio reviews	At least one portfolio review/year
203.3.8 Operating Unit Annual Operational Plan and Annual Performance Report	Purpose of performance reports, data quality, and environmental compliance reports	A yearly operational plan and a yearly performance report, and a report on compliance with environmental requirements
203.3.9 Reporting Requirements for Projects Not Managed by Country-Based USDH Staff	Specifications of reporting requirements for USAID-funded projects not managed by country-based USDH staff	None
203.3.10 Intensive Program Reviews	Examination of the progress of AOs	At least one intensive review every three years
203.3.11 Assistance Objective Close Out Reports	Suggested contents of close out reports	A brief close out report for each AO on either the completion or termination of the AO
203.3.12 Development Experience Clearinghouse	A list of documents subject to submission	Information must be shared via the DEC

**The “USAID Evaluation Policy”** of 2011 provides general principles that evaluations should follow.<sup>15)</sup> According to the evaluation policy, the two purposes of evaluation are “accountability” and “learning.” The role of evaluation is to inform USAID on the decision-making process by providing sound and transparent evidence. For these goals, the policy recommends the following:

- 1) Well-designed evaluations and well-functioning feedback loops for sharing information are essential components;

<sup>15)</sup> USAID (2011), “USAID Evaluation Policy”



- 2) Organizational efforts, especially those of senior managers to encourage the active involvement of all staff members in the evaluation process, are crucial;
- 3) Sufficient resources must be dedicated to evaluation; and
- 4) The methods of evaluation should be scientific<sup>iv)</sup> and appropriately address the evaluation questions.

At a glance, the policy does not look new. However, all of the recommendations come from the history of evaluation in USAID. As previously described, evaluation had not always functioned as it should have. It was often pushed away and confused with performance monitoring. The importance of evaluation in aid effectiveness was ignored for a long time. The USAID Evaluation Policy aims to communicate the basic principles of evaluation.

## Guidelines and Policy: Overview

Based on a review of three documents above, the following overview has been constructed.

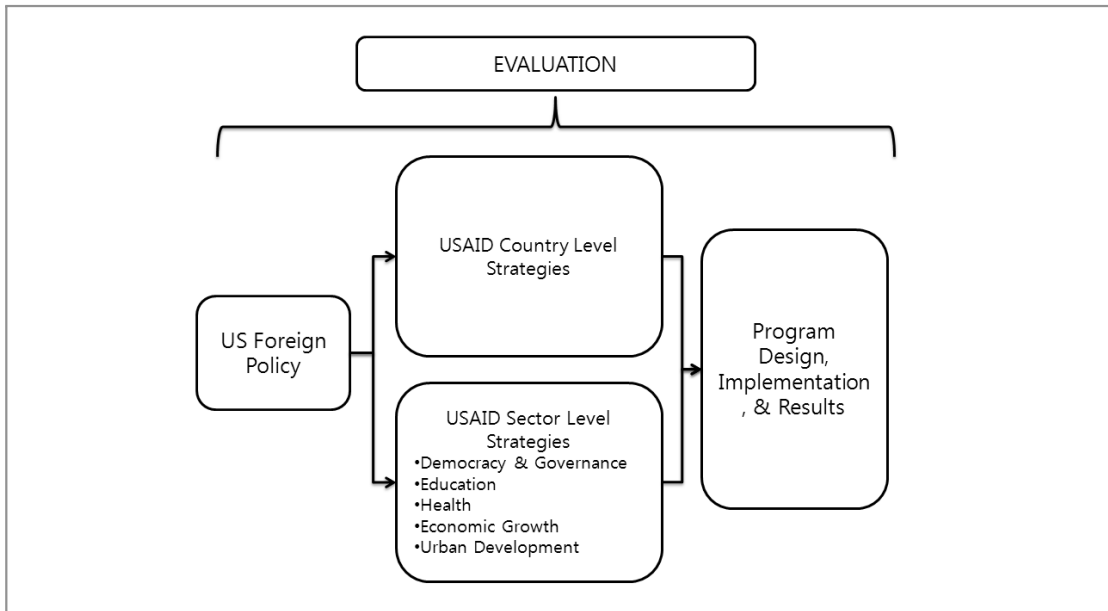
### The Definitions of Evaluation

USAID defines evaluation as “the systematic collection and analysis of information about the characteristics and outcomes of programs and projects as a basis for judgments, to improve effectiveness, and/or inform decisions about current and future programming.”<sup>16)</sup> This definition is consistent throughout all three documents. Evaluation is carried out for all levels of operation in USAID: policy, country/sector level strategies, and programs. The scope and primary purpose of evaluation differs at each level. For instance, performance evaluations may be more appealing to program implementation partners, while USAID’s upper management staff is more concerned with impact evaluations of programs and policies.

---

16) USAID (2011), “USAID Evaluation Policy”

〈Figure 2〉 Evaluation in USAID



There are several types of evaluation to choose from. Box 1 contains definitions of terms related to evaluation used by USAID.

〈Box 1〉 Types of Evaluation

- Impact evaluations measure the change in a development outcome that is attributable to a defined intervention; impact evaluations are based on models of cause and effect and require a credible and rigorously defined counterfactual to control for factors other than the intervention that might account for the observed change. Impact evaluations in which comparisons are made between beneficiaries that are randomly assigned to either a treatment or a control group provide the strongest evidence of a relationship between the intervention under study and the outcome measured.
- Performance evaluations focus on descriptive and normative questions: what a particular project or program has achieved (either at an intermediate point in execution or at the conclusion of an implementation period); how it is being implemented; how it is perceived and valued; whether expected results are occurring; and other questions that are pertinent to program design, management and operational decision making. Performance evaluations often incorporate before–after comparisons, but generally lack a rigorously defined counterfactual.

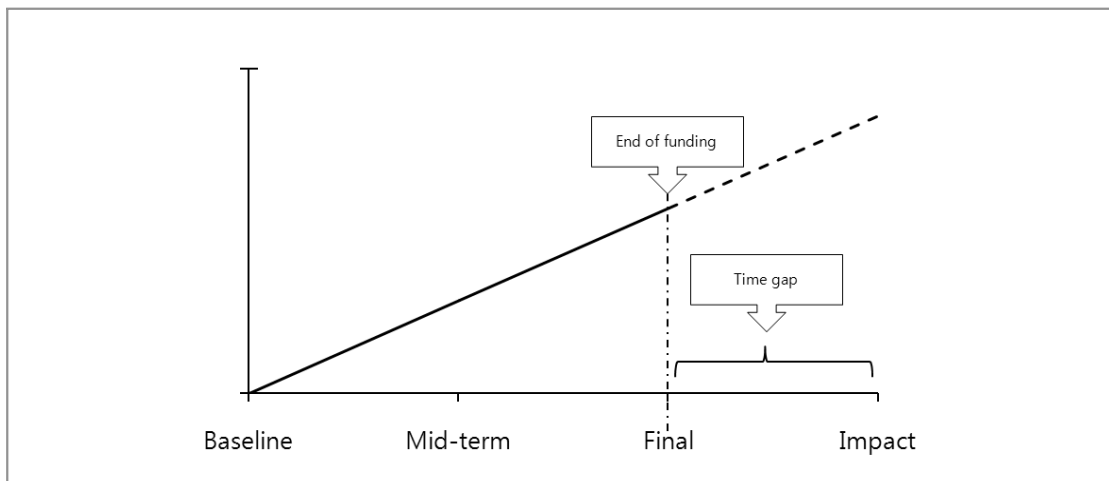
- Performance monitoring of changes in performance indicators reveals whether desired results are occurring and whether implementation is on track. In general, the results measured are the direct and near-term consequences of project activities.

Source: USAID (2011), Evaluation Policy, p.2

Defining the terms is an important step in evaluation. It puts all stakeholders of evaluation - managers, evaluation teams, other related staff members, and implementing partners - on the same page and ensures that USAID obtains accurate results.

Impact evaluation and performance evaluation differ in both timing and purpose. An impact evaluation is an evaluation for obtaining long-term results of a program and policies. It is usually done after the program has ended and there is a time gap between the end of the funding/activity and the evaluation. A performance evaluation is done during the funding/activity cycle and focuses on the short-term results of a program.

〈Figure 3〉 Evaluation Time Line



More importantly, the policy distinguishes performance evaluation from the performance monitoring. The confusion of these two concepts was identified as a reason why evaluation degraded both in quantity and quality in the past.<sup>17),18),19)</sup> Monitoring is continuous,

17) USAID (2011), "USAID Evaluation Policy"

18) Cynthia Clapp-Wincek and Richard Blue (2001), "Evaluation of Recent USAID Evaluation Experience," USAID PPC/CDIE, document PN-ACG-632.

19) Janice M. Weber (2004), "An Evaluation of USAID's Evaluation Function: Recommendations for Reinigorating

day-by-day activity done by implementation staff themselves. The measurements are taken and compared against predetermined target levels. It functions as a tool for better implementation of a program. The data collected usually only answers “what” and “when” something happened. An evaluation is more of an event or episode. The purpose of an evaluation is to answer “why” and “how.” The whole picture of a program emerges only when monitoring and evaluation are balanced.

## Communicating the Results

The guidelines make it clear to whom and what to report at the end of each evaluation. The procedure of dissemination of findings of an evaluation is laid out consistently through all three documents, specifying responsible parties and methods. This comes from past experience. USAID was “born decentralized”; that is, major policies guided the directions of operations, and the local USAID offices were fully in charge and responsible for their own field operations at the field level. Evaluations were also decentralized; performance monitoring and evaluations were mostly the responsibility of the implementing partner.<sup>20)</sup> This resulted in an inconsistency in the quality of evaluations, and there was also a lack of information sharing. Implementing partners were not necessarily willing to share their evaluation findings with other USAID divisions with which they had no direct relationship. Therefore, many of the evaluation reports did not make it to the Washington office. Currently, evaluation is under the guidance of a central evaluation unit and it is mandatory to submit the evaluation reports to the Development Experience Clearinghouse.

## Evaluation Criteria

USAID did not necessarily follow the DAC recommended evaluation criteria in the past. Accepting a results-based management (RBM) approach as the agency’s management principle, USAID considered evaluation as a tool for obtaining specific answers rather than an assessment tool for performance and the evaluation criteria were defined on what the evaluation needed to answer.<sup>21)</sup> Current criteria are very similar to DAC’s recommended five

---

the Evaluation Culture within the Agency”

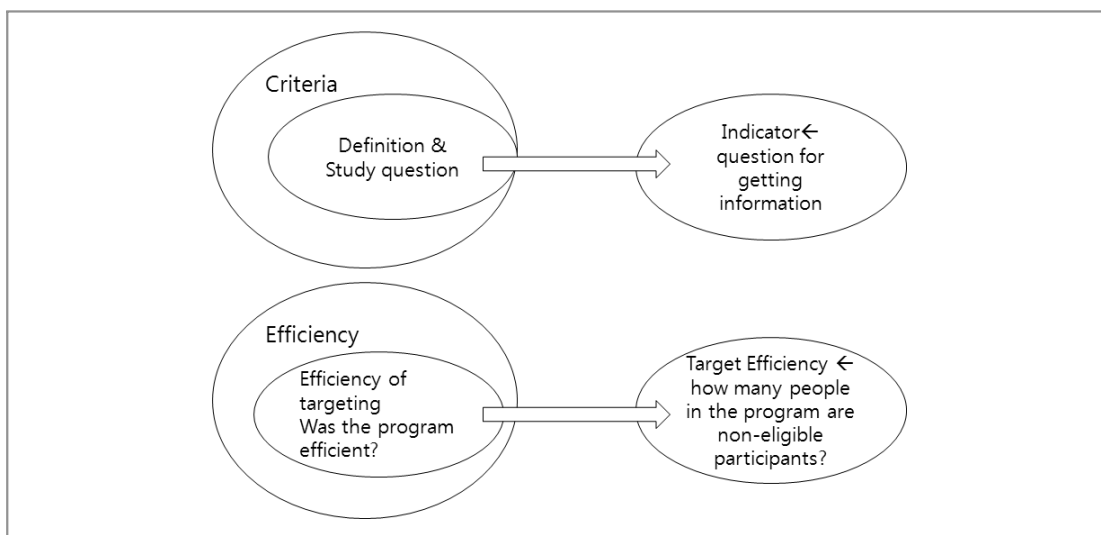
20) USAID (2009), “Trends in International Development Evaluation Theory, Policy and Practices,” Washington, D.C., USA

criteria of evaluation; the “Evaluation Guidelines for Foreign Assistance” recommends choosing at least one among these five criteria (relevance, effectiveness, efficiency, impact and sustainability) to be examined based on the objective of the evaluation.

Selecting and defining the criteria provides a basis for forming “study questions” and eventually for choosing indicators. The guidelines recommend that all evaluations set “study questions” (e.g., what has the project achieved?). Once the study question is defined, the evaluation should be designed in a manner that would be the most appropriate to answer that question.

As an example, if “efficiency” is needed to be examined, efficiency in this program is defined as the cost efficiency in program delivery. Based on this, a study question, “Was the program efficient in delivering aid to the proper beneficiaries?” is developed. There are various indicators that can be used to examine the efficiency of program delivery. Various related indicators are examined, and the indicator for “target efficiency” is chosen. Target efficiency can be measured by simply asking program implementers, “How many people in the program are non-eligible participants?” or participants themselves whether they meet the participant selection criteria.

〈Figure 4〉 Criteria, Study Question, and Indicator



21) Ryoh Sasaki (2006), “A Review of the History and the Current Practice of Aid Evaluation,” Journal of Multidisciplinary Evaluation, Number 5, pp.55–88

As specified in guidelines, most evaluation reports of USAID programs are formatted in the “question” and “answer” format.<sup>v)</sup>

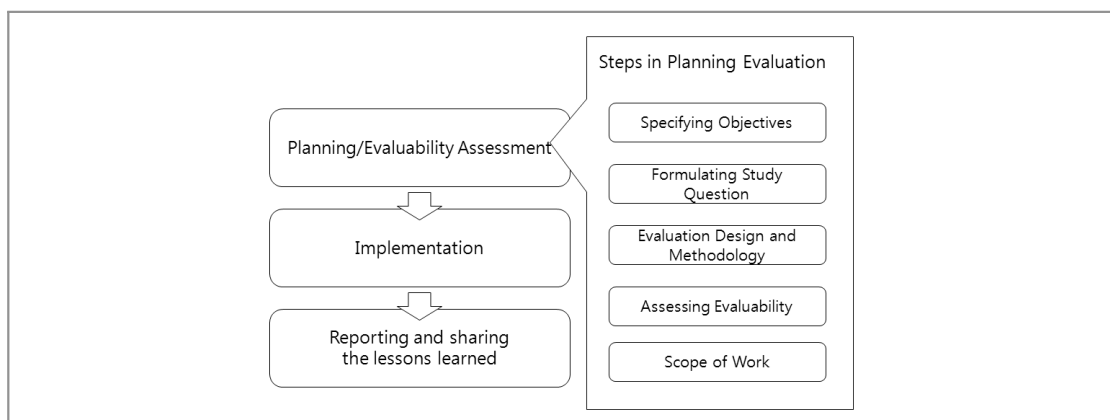
## Evaluation in USAID: In Practice

In this section, three evaluations of USAID are used as cases for detailed examination on what is implemented during an evaluation.

### Evaluability Assessment: The USAID/Brazil Micro and Small Enterprise Trade-Led Growth Program Evaluation

The first question that should be asked in the evaluation process is whether to carry out an evaluation or not. An evaluation, even the simplest one, will take time and resources. When the program is not ready to be evaluated, the carrying out of an evaluation puts not only an unnecessary burden on resources but also may weaken the program impact itself by rushing judgment. To avoid such consequences, evaluability assessments are recommended. An evaluability assessment is the examination of evaluation feasibility. The process provides in-depth information and insights over the program. This is also an opportunity to form a consensus among the program stakeholders on the evaluation. Three major questions should be answered by the end of the evaluability assessment: the plausibility (in terms of the impact size), the feasibility (in terms of resources and time given to the evaluation) of measuring the impact, and whether or not the impact evaluation would be useful.<sup>22)</sup>

〈Figure 5〉 Evaluation Procedure



22) Elizabeth Dunn (2008), “Planning for Cost-Effective Evaluation with Evaluability Assessment”

One example of an evaluability assessment in action can be found in the evaluability assessment of the USAID/Brazil Micro and Small Enterprise Trade-Led Growth Program Report, which examined the feasibility of the original design of an impact evaluation against the implementation of a program.<sup>23)</sup> The program itself was a small pilot program implemented in northeast part of Brazil; the goal of this program was a trial for small Brazilian firms to access the international market. The program intended to help the firms meet market requirements and to create exporter-importer linkages between the firms. If the program worked, these firms were expected to increase purchases from local microenterprises. After examining the implementation of the program and the data accumulated during the monitoring, the team concluded that a full extensive impact evaluation would not be recommended due to: a) the fact that the scale and the duration of program were not sufficient to make a meaningful impact, since it only lasted two years and three months, and b) resource restrictions; it was not plausible to carry out a full-blown impact evaluation. In 2007, an impact assessment was carried out as recommended by an evaluability assessment and various methodological concerns were addressed along with presenting the impact of the program.<sup>24)</sup>

#### Program Performance Evaluation: USAID/Kosovo SME and Agricultural/Agribusiness Programs<sup>25)</sup>

Between 1999 and 2004, USAID Kosovo Office implemented a complex program, consisting of 10 projects to enhance small enterprises, the agricultural market and related industries, as well as agricultural productivity. In order to examine the performance, an evaluation was carried out by an internal evaluation team from USAID in 2004.

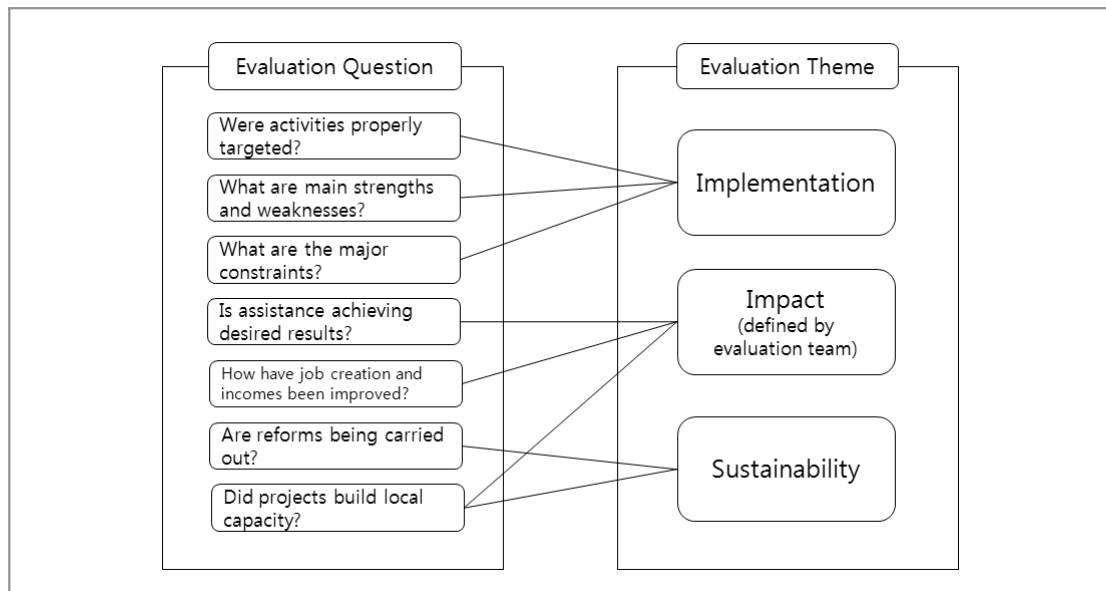
This evaluation was carried out when the emphasis was still on the monitoring and treating evaluation as a means of obtaining answers for specific questions. The Kosovo evaluation team had classified evaluation questions into three categories: implementation, impact, and sustainability. In addition, specific questions were asked in order to get information on each category. For each project, the evaluation was carried out to answer those questions.

23) Don Snodgrass *et al.* (2006), "Evaluability Assessment of the USAID/Brazil Micro and Small Enterprise Trade-Led Growth Program"

24) Bonnie Brusky and Joana Monteiro (2008), "Assessing the Impact of the Micro and Small Enterprise Trade-Led Growth Project of USAID/Brazil"

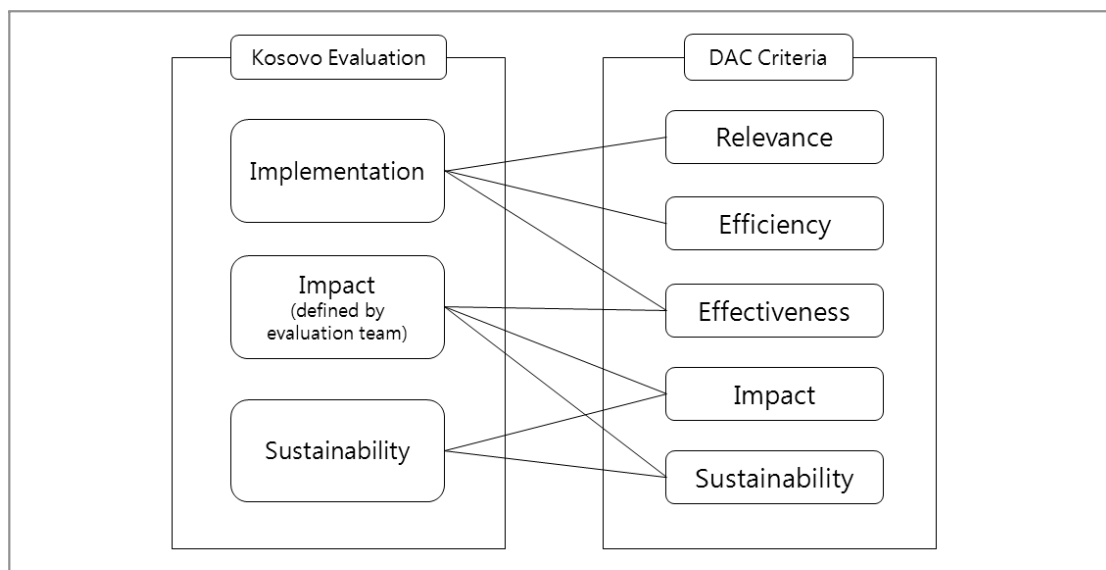
25) USAID (2004), "Evaluation of the USAID/Kosovo SME and Agricultural/Agribusiness Programs"

〈Figure 6〉 Evaluation Questions & Criteria of Kosovo SME and Agricultural/Agribusiness Programs Evaluation



Even though the term itself overlaps somewhat with DAC evaluation criteria, the definition is different. For instance, the impact of the Kosovo evaluation appeared to encompass three of DAC evaluation criteria: effectiveness, impact and sustainability. Even though the term's definition differs, the Kosovo evaluation dealt with most of the DAC criteria.

〈Figure 7〉 Comparison of Evaluation Categories: Kosovo Evaluation vs. DAC Guidelines





### Program Impact Evaluation: Rebuilding Agricultural Markets Program, Afghanistan

As previously described, a team of external evaluators performed an impact evaluation on the Rebuilding Agricultural Market Program (RAMP), which ended in 2006. Begun in 2003, RAMP was a program designed to rehabilitate and to revitalize the agricultural sector, mainly by providing essential infrastructure such as irrigation and roads.

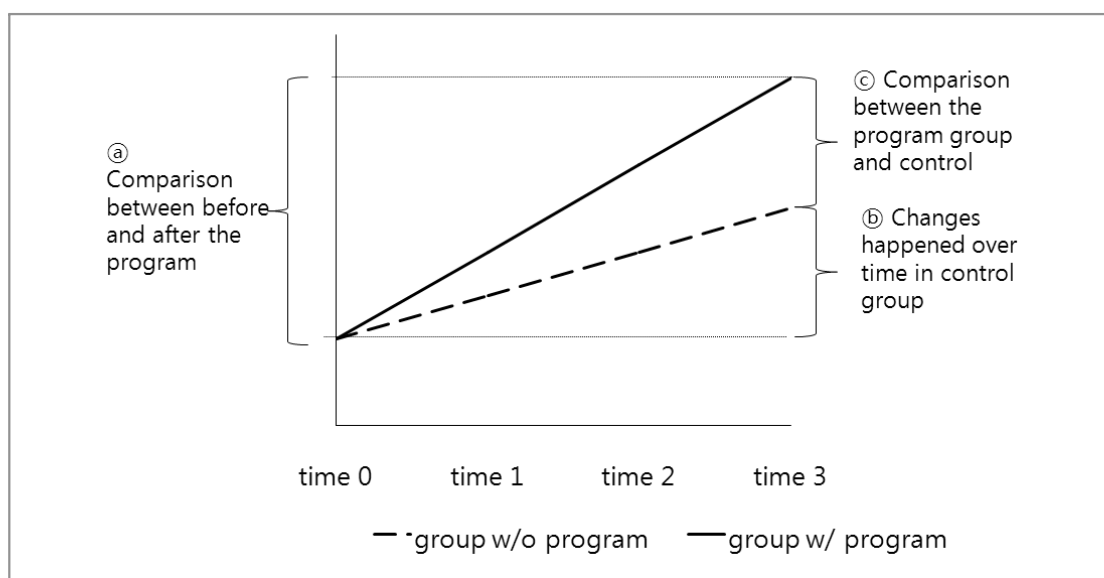
As in the Kosovo evaluation above, this evaluation was carried out to answer specific evaluation questions, such as the extent of achievement and appropriateness of the strategy used in the program to address problems.

#### 〈Box 2〉 Study Questions in the Impact Evaluation of RAMP<sup>vi)</sup>

- A. Based upon program activities, program outputs and program outcomes, assess the extent to which its overall objective was achieved.
  - 1. Was the program successful in addressing the needs of Afghanistan's agricultural sector and its target beneficiaries?
  - 2. What difference did the program interventions have in the provinces visited?
  - 3. Did RAMP effectively support USAID/Afghanistan's agricultural objectives?
  - 4. To what extent did RAMP support other programs/projects funded by USAID?
- B. To what extent has RAMP successfully trained and involved the Ministry of Agriculture staff in its activity planning and implementation?
- C. Was the general strategy of demand-driven approach a sound strategy considering the economic and political situation at the time of program implementation?
- D. The implementation of activities under market development and infrastructure was through implementation partners by issuing job orders for specific area activities. Was this implementation arrangement effective in terms of resource allocation and monitoring the progress of the program in achieving its goal?
- E. Chemonics submitted an economic assessment of the impact of the program
  - 1. Review the soundness of the methodology and approach used in quantifying the program's impact
  - 2. Review the validity of major results findings, especially the overall impact on the agricultural sector of \$1.7 billion and the 823 kilometers of irrigation canals rehabilitated
- F. Review the monitoring system used to evaluate the impacts of the activities. How effective was the monitoring system?

This evaluation is an example of idealism-meeting-the-field in evaluation design. Though named as an “impact evaluation,” it chose to examine only trends over time without a control group. This is the weakest of all impact evaluation designs. An impact evaluation should be able to define the size of impact attributable to the program. Therefore, impact evaluations tend to use study designs comparing situations before and after the program, and also against comparison groups where the program was not implemented during the period under evaluation.

〈Figure 8〉 Experimental Design in Impact Evaluation



In clinical trials, the gold standard is a randomized, double-blind study with a control group. The participants are assigned into either a treatment group or a control group at random and the group members and staff who are in direct contact with study participants should not know whether the treatment is a true treatment or placebo. This strict design provides the most statistically sound results, but it is also next to impossible to be used in development aid programs. Instead, a program impact evaluation often uses a design known as a “quasi-experimental design”; that is, the program does not use the random assignment of participants. By comparing the pre- and post-program values of certain indicators (e.g., farmers’ incomes), the changes in the indicator related to the impact of program can be identified (Figure 8, (a)). However, without a control group, the difference is not clear as to whether an impact is due to other factors (e.g., improvement in security in Afghanistan) or

due to the program service, since the changes can happen even without the program (in Figure 8, ㉞). Having a control group solves this problem. By comparing changes in program groups against those of control groups, the attributable impact can be identified (in Figure 8, ㉟).

The use of experimental design appeared very early in USAID evaluation. An evaluation handbook published in 1970 contains diagrams and guidelines for such a design. However, often this ideal design does not survive the reality in the field.

Considering the Afghanistan situation, application of a true experimental design is not realistic. That does not mean that impact evaluations should not be carried out. The impact of RAMP was important for program implementers and USAID management. Instead of abandoning the possibility, the evaluation team provided the best possible answers to evaluation questions while acknowledging the limitations. It was also possible due to strong and extensive monitoring of data collected during the course of implementation. The evaluation provided valuable information for the program team. For instance, the impact evaluation found that the farmers often did not benefit from the program and the size of economic impact was much smaller than what past reports had suggested. It might seem to be a harsh criticism of the program. However, the suggestions may give more credibility to RAMP impacts in the long run.

### III. Conclusion: The Lessons Learned from USAID Evaluation Experience

This article reviews the overall policy and procedures of USAID and examined three cases of evaluation in order to provide insights for successful evaluation. Some practices of USAID are not feasible to apply to KOICA. For instance, USAID uses the scope of work as an effective communication tool for revealing the purpose of an evaluation, and also a means to control the quality of the evaluation. The quality assurance criteria of evaluation reports clearly specify that “evaluation reports shall address all evaluation questions included in the scope of work.”<sup>26)</sup> However, this may not be applicable to emerging donor agencies, where the

26) USAID (2011), “USAID Evaluation Policy,” <http://www.usaid.gov/evaluation>, (accessed Mar. 10, 2011)

evaluation division has not yet accumulated a pool of experienced experts, and therefore may not be able to cover all necessary details. On the other hand, policies such as giving emphasis on evaluation and nurturing an evaluation-friendly culture should be adopted in the early stages. USAID invented and accumulated evaluation but drifted away from it, resulting in the loss of quality and quantity of evaluations, and eventually the effectiveness of development aid. Also, procedures such as evaluability assessments can be used to prioritize resource allocation for evaluation. The following are suggestions made based on the review.

An evaluation reflects and stores lessons learned during a program. It is an important vehicle for organizational learning. Decentralized organizations, such as USAID, need efficient systems for collecting evaluations, and also clear guidance on the responsibilities of all parties involved in the evaluations.

An agency should acknowledge the importance of evaluation. Evaluation should be supported by organizational policy, otherwise the quality and quantity of evaluations can suffer and eventually the efficiency and effectiveness of programs may deteriorate. On the same note, evaluation and also monitoring of the activities should be included in program/policy design. Evaluation should be considered as an essential part of all activities, including strategy forming and designing and implementing programs and policies. Evaluation, however, should not be done for the sake of evaluation. In order to avoid an unnecessary burden due to an evaluation, an evaluability assessment should be done prior to the evaluation and necessary adjustments should be made.

Having a clear goal for evaluation is essential in successful evaluation. All study questions should be laid out first and communicated clearly with the evaluation team. Also, the evaluation questions, definitions of terms and eventually what is expected to be done by the evaluation team should be shared among all stakeholders of evaluation.

With the new evaluation policy, hopes are up for USAID. However, it is too early to examine whether the new USAID evaluation policy is making an actual difference in terms of evaluation. It would be exciting to see how the new policies and guidelines, the result of a long struggle to revive evaluation and aid effectiveness at USAID, affect evaluations on the ground.

## References

- Brusky, Bonnie and Joana Monteiro (2008), “Assessing the Impact of the Micro and Small Enterprise Trade-Led Growth Project of USAID/Brazil”
- Clapp-Wincek, Cynthia and Blue, Richard (2001), “Evaluation of Recent USAID Evaluation Experience,” USAID PPC/CDIE, document PN-ACG-632
- Corneille, Faith (2001), “Entangled Accountability in Development Agencies: The Case of USAID”
- Dunn, Elizabeth (2008), “Planning for Cost-Effective Evaluation With Evaluability Assessment”
- Levine, Ruth (2011), “USAID’s Evaluation Policy: Setting the Standard,” <http://blog.usaid.gov/2011/01/usaid-evaluation-policy-setting-the-standard>, accessed May 15, 2011
- OECD DAC (2006), “The United States DAC Peer Review”
- Office of the Director of U.S. Foreign Assistance (2009), “Evaluation Guidelines for Foreign Assistance”
- Sasaki, Ryoh (2006), “A Review of the History and the Current Practice of Aid Evaluation” *Journal of Multidisciplinary Evaluation*, Number 5, pp.55–88
- Snodgrass, Don et al. (2006), “Evaluability Assessment of the USAID/Brazil Micro and Small Enterprise Trade-Led Growth Program”
- USAID (2004), “Evaluation of the USAID/Kosovo SME and Agricultural/Agribusiness Programs”
- USAID (2005), “Note from Natsios” *Frontline*, March, p.3
- USAID (2007), “Impact Evaluation: Rebuilding Agricultural Markets Program,” Washington, D.C., USA
- USAID (2009), “Trends in International Development Evaluation Theory, Policy and Practices,” Washington, D.C., USA
- USAID (2010), “ADS Chapter 203: Assessing and Learning”
- USAID (2011), “USAID Evaluation Policy”
- Weber, Janice (2004), “An Evaluation of USAID’s Evaluation Function: Recommendations for Reinvigorating the Evaluation Culture within the Agency,” USAID
- White, Howard (2005), “Challenges in Evaluating Development Effectiveness,” *Institution of Development Studies*, Brighton, England

## Note

- i) However, due to the special circumstances, operations such as natural disaster response are exempted from this obligation.
- ii) The DEC is USAID's online repository for all documents and records on various aspects of USAID, which was established in 1975. All evaluation reports (either the final report when the evaluation is completed or the final draft when the evaluation is not completed) must be submitted either in electronic form or in paper copy. Documents and materials can be found and downloaded at <http://dec.usaid.gov/>.
- iii) The Automated Directives System (ADS) is a standardized system of USAID containing: (1) USAID internal mandatory guidance, including policy directives and required procedures; (2) external mandatory guidance applicable to USAID; and (3) non-mandatory guidance. The purpose of this system is to provide aid to employees and partners in interpreting and applying internal and external mandatory guidance.
- iv) This is with regard to the reproducibility issues, meaning the methods are expected to yield relatively similar findings when other qualified evaluators apply them.
- v) See Box 2 for an example of the “Q&A” format.
- vi) These questions were quoted directly from the impact evaluation report.