

Trends in Development Assistance—Series 5

Evaluating Development Assistance: A Japanese Perspective



Edited by
MINATO Naonobu
FUJITA Nobuko

FASiD

Foundation for Advanced Studies on International Development
International Development Research Institute (IDRI)

Evaluating Development Assistance: A Japanese Perspective

The Foundation for Advanced Studies on International Development (FASID) was established in April 1990. FASID and its affiliate, International Development Research Institute (IDRI), conduct research, facilitate interaction among researchers and practitioners, and offer training programs to development specialists. These activities are aimed for improvement in the quality of development programs and policies.

This publication is financially supported by the Ministry of Foreign Affairs of Japan.

Copyright© 2009 by FASID

Published in 2009 in Japan by the Foundation for Advanced Studies on International Development, 1-6-17 Kudan-minami, Chiyoda-ku, Tokyo 102-0074, Japan

e-mail: pub@fasid.or.jp

URL: <http://www.fasid.or.jp>

Preface

Currently, activities of FASID International Development Research Institute (IDRI) are centered around three pillars: (i) Researching aid strategies as an ODA policy think tank; (ii) Functioning as a knowledge hub that offers a forum for debates between unconfined and diverse viewpoints; and (iii) Research on and practice of ODA evaluation. This publication — the fifth in the Trends in Development Assistance Series — focuses on evaluation, the third pillar, and provides front-line reports on recent trends and information of interest to domestic and foreign practitioners, policy makers, and educational and research institutions.

Evaluation is absolutely necessary for improving the quality and transparency of development assistance. Research and discussions on evaluation are advancing at academic conferences and donor meetings and revealing the diverse aspects of evaluation. Every year, new research findings on evaluation are reported at the American Evaluation Association, Japan Evaluation Society and other academic associations. At international forums, topics such as the implementation status of the Paris Declaration on Aid Effectiveness by the OECD/DAC Network on Development Evaluation, and the pros and cons of impact evaluations by The World Bank and others are being actively discussed and debated. Discussions are also taking place on diverse topics such as the independence and ethics of evaluation, as well as how to improve developing countries' evaluation capacity.

In Japan, set against the background of tight fiscal conditions and ongoing administrative reforms, demand for better accountability to taxpayers is increasing together with the demand to improve the effectiveness of ODA. The role of evaluation in gaining the people's confidence in ODA is growing in importance. While some "evaluation fatigue" can be observed in the area of domestic administrative evaluation, in the area of ODA evaluation we expect the merger of JICA and JBIC will allow an integration of aid modalities, leading to new developments in the framework of evaluation activities. In March 2009, the governments of Japan and Singapore held an ODA Evaluation Workshop in Singapore. Representatives of 20 aid recipient countries in Asia and aid organizations participated in the Workshop and engaged in spirited discussions on improving policy, project-level, and joint evaluations. In light of these trends in development aid evaluations, this publication describes and analyzes evaluation's main functions and challenges.

This publication contains articles written by Hiromitsu MUTA

(Executive Vice President for Finance, Tokyo Institute of Technology) and Yuriko MINAMOTO (Associate Professor, Meiji University), Kiyoshi YAMAYA (Professor, Doshisha University), Takako HARAGUCHI and Keishi MIYAZAKI (consultants), Ryokichi HIRONO (Professor Emeritus, Seikei University), and Michael BAMBERGER (consultant). We would like to express our sincere gratitude to these contributors who have been researching and practicing evaluation for many years in their respective areas of specialization. Each chapter of this publication is based on the opinions of each author and does not represent the opinions of the organizations to which the author belongs. Authors were associated with the organizations mentioned here at the time of writing.

We would also like to convey our appreciation to Mr. Hajime SATO for translating chapters 1 and 2, and Mr. Paul CONSALVI for proofreading. In addition, we extend our appreciation to Akiko TSUYUKI and Nao TAKAYAMA for editorial assistance. It would be our pleasure if this publication can contribute to improving the quality of Japan's ODA through evaluation.

March 2009

Naonobu MINATO
Nobuko FUJITA
Editors

Trends in Development Assistance –Series 5– Evaluating Development Assistance: A Japanese Perspective

Preface

MINATO, Naonobu

FUJITA, Nobuko

List of Tables, Figures and Boxes

Abbreviations and Acronyms

Chapter 1

Development Assistance Evaluation in Japan: Challenges and Outlook

MUTA, Hiromitsu

MINAMOTO, Yuriko1

Chapter 2

ODA Evaluation and Policy Evaluation: Status of Accountability and
Transparency in Japan

YAMAYA, Kiyoshi28

Chapter 3

Evaluation Capacity Development: A Practical Approach to Assistance

HARAGUCHI, Takako

MIYAZAKI, Keishi53

Chapter 4

Evaluation Capacity Development in the Asia-Pacific Region: A Proposal
for an Asia-Pacific Evaluation Association Network (APEA NET)

HIRONO, Ryokichi96

Chapter 5

Institutionalizing Impact Evaluation Systems in Developing Countries:
Challenges and Opportunities for ODA Agencies

BAMBERGER, Michael127

Editor's Notes

FUJITA, Nobuko168

About the Authors173

List of Tables, Figures, and Boxes

Tables

Table 1-1	Main recommendations of the <i>Report on Reform of Japan's ODA Evaluation System</i> (March 2000)	4
Table 1-2	Evaluations conducted by MOFA, JICA, and JBIC (FY2005).....	6
Table 1-3	Main issues of Japan's ODA evaluation	23
Table 2-1	Evaluations requested by the Ministry of Internal Affairs and Communications to the Ministry of Foreign Affairs (FY2002-04)	32
Table 2-2	Various type of ODA Evaluation	47
Table 3-1	JBIC's ECD Assistance Model for ODA Loan Projects.....	55
Table 3-2	Outlines of ODA Loan Project Evaluation Seminars 2002 and 2007	56
Table 3-3	Roles of concerned agencies in conventional and joint evaluations	61
Table 3-4	Key Milestones for the Development of Legal Framework regarding M&E of ODA in Vietnam	76
Table 3-5	Outline of Decision No. 1248/2007/QD-BKH of MPI	77
Table 3-6	Major Outputs of VAMESP II	79
Annex Table 3-1-1	Details of Evaluation Training ODA Loan Projects (Example of 2007 Seminar)	88
Annex Table 3-1-2	Details of Evaluation Systems Workshop Modules (Example of 2007 Seminar)	88
Annex Table 3-1-3	Problems and Measures on Evaluation Systems in Countries Participating in ODA Loan Project Evaluation Seminars (2004, 2005 and 2006)	89
Annex Table 3-2-1	Japanese ODA Projects under the Red River Transport Development Program (1994-2004)	90
Annex Table 3-2-2	Comparison of Two Joint Evaluations in 2005 and 2007	92
Table 4-1	Millennium Development Goals So Far Achieved in All Regions	119
Table 4-2	Socio-Economic Indicators of Development in All Regions ...	119
Table 4-3	Unemployment, Income and Gender Inequality and Access to Electricity in Asia-Pacific Countries	120
Table 4-4	Investment, Trade, Aid and Finance in All Regions	120
Table 4-5	Governance in Asia-Pacific Countries, 2000	121

Table 4-6	Governance Indicators in Asia-Pacific Region, 2007	121
Table 5-1	Types of project or program evaluation used at different stages of the project cycle	135
Table 5-2	Widely used impact evaluation designs	138
Table 5-3	Key steps for institutionalizing impact evaluation at the national and sector levels	146
Table 5-4	Incentives for IE—Some Rewards [“Carrots”], Sanctions [“Sticks”], and Positive Messages from Important People [“Sermons”]	150
Table 5-5	IE Skills and Understanding Required by Different Stakeholder Groups	153
Table 5-6	Examples of the Kinds of Influence IEs Can Have	158

Figures

Fig. 1-1	Organizations that conduct ODA evaluations and their evaluation subjects	5
Fig. 2-1	Various Evaluation Issues	30
Fig. 2-2	Areas where evaluation and similar “evaluation” type activities take place.....	35
Fig. 2-3	Division of labor system for ODA evaluations	42
Fig. 3-1	Factors of Learning through ODA Loan Project Evaluation Seminars	63
Fig. 5-1	Three Pathways for the Evolution of Institutionalized IE Systems.....	141

Boxes

Box 5-1	The danger of over-estimating project impact when the evaluation does not collect information on comparison groups who have not benefited from the project intervention	131
Box 5-2	Two alternative definitions of impact evaluation [IE]	134
Box 5-3	Colombia: Moving from the Ad Hoc Commissioning of IE by the Ministry of Planning and Sector Ministries toward Integrating IE into the National M&E System (SINERGIA) ...	140
Box 5-4	Mexico: Moving from an Evaluation System Developed in One Sector toward a National Evaluation System (SEDESOL)	142
Box 5-5	Africa Impact Evaluation Initiative	144
Box 5-6	Chile: Rigorous IEs Introduced as Part of an integrated Whole-of-Government M&E System	145

Abbreviations and Acronyms

ADB	Asian Development Bank
AfrEA	African Evaluation Association
AGM	Annual General Meetings
AIDS	Acquired Immune Deficiency Syndrome
AIM	Africa Impact Evaluation Initiative
A/P	Asia-Pacific
APEA NET	Asia-Pacific Evaluation Association Network
AusAID	Australian Agency for International Development
CCBP	Comprehensive Capacity Development Program
CDF	Comprehensive Development Framework
CDOPP	Capacity Development of ODA Project Planning (JICA)
CIDA	Canadian International Development Agency
COE&GP	Center of Excellence and Good Practice
DAC	Development Assistance Committee
ECD	Evaluation Capacity Development
GDP	Gross Domestic Product
GNI	Gross National Income
GPEA	Government Policy Evaluations Act
GSO	General Statistical Office
HDR	Human Development Report
HICs	High Income Countries
HIV	Human Immunodeficiency Virus
IAs	Incorporated Administrative Agencies
IDEAS	International Development Evaluation Associations
IE	Impact Evaluation(s)
IEG	Independent Evaluation Group
IOCE	International Organization for Cooperation in Evaluation
JBIC	Japan Bank for International Cooperation
JES	Japan Evaluation Society
JICA	Japan International Cooperation Agency
JOCV	Japan Overseas Cooperation Volunteers
LICs	Low Income Countries
M&E	1) Measurement and Evaluation 2) Monitoring and Evaluation
MAFF	Ministry of Agriculture, Forestry and Fisheries (Japan)
MDGs	Millennium Development Goals

MES	Malaysian Evaluation Society
MEXT	Ministry of Education, Culture, Sports, Science and Technology (Japan)
MIC	Ministry of Internal Affairs and Communications (Japan)
MICs	Middle Income Countries
MLIT	Ministry of Land, Infrastructure, Transport and Tourism (Japan)
MOE	Ministry of the Environment (Japan)
MOF	Ministry of Finance (Japan)
MOU	Memorandum of Understanding
MPI	Ministry of Planning and Investment (Vietnam)
NGO	Non-Governmental Organization
NIMES	National Integrated M&E System
NONIE	Network of Networks for Impact Evaluation
NPM	New Public Management
O&M	Operation and Maintenance
ODA	Official Development Assistance
OECD	Organization for Economic Cooperation and Development
PAP	People's Action Party
PCM	Project Cycle Management
PDM	Project Design Matrix
PMU	Project Management Unit
PPP	Purchasing Power Parity
PRSPs	Poverty Reduction Strategy Papers
R&D	Research and Development
REA	Regional Evaluation Association
SIDA	Swedish International Development Agency
SLEvA	Sri Lanka Evaluation Association
SSIs	Semi-Structured Interviews
UMNO	United Malay National Organization
UNDP	United Nations Development Programme
UNICEF	United Nations Children's Fund
VAMESP	Vietnam Australia Monitoring and Evaluation Strengthening Project
VDGs	Vietnam Development Goals
WDR	World Development Report

Development Assistance Evaluation in Japan: Challenges and Outlook

Hiromitsu Muta, Yuriko Minamoto

1. Introduction

Since the passage of the Government Policy Evaluations Act of 2001, all Japanese ministries have been evaluating policies, measures, administrative affairs, and executive agencies programs (policy evaluations). Prior to passing the 2001 Act, the evaluation function was understood to be necessary and related ministries had been conducting evaluations for public works projects, research and development, Official Development Assistance and other areas requiring large amounts of resources. Particularly in regards to Official Development Assistance (ODA), Japan recognized the importance of evaluations early and has endeavored to establish evaluation systems for ODA. As early as 1975, the Japan Bank for International Cooperation (JBIC, formerly the Overseas Economic Cooperation Fund) began to conduct ex-post evaluations, and from 1991 started publishing *Ex-Post Evaluation Reports on ODA Loan Projects* (the Ministry of Foreign Affairs, Economic Cooperation Bureau 1999). In 1981, MOFA created the Economic Cooperation Evaluation Committee within the Economic Cooperation Bureau and began its own ex-post evaluations. Since 1982, MOFA has been publishing its *Annual Evaluation Reports on Japan's Economic Cooperation*. In 1984, the Research and Programming Division was created to administer evaluations, and in 1990, its evaluation group was split-off to become the ODA Evaluation Division. In 1981 the Japan International Cooperation Agency (JICA), also created its own Evaluation Study Committee and the following year began ex-post evaluations. Since 1995, it has also been publishing *Annual Evaluation Reports*. The evaluation of ODA activities has been demanded for

a long time because the public cannot easily see ODA activities with their own eyes. Early on in the history of ODA, evaluations based on international standards have been conducted on ODA activities because they take place overseas, can involve joint work, and are subject to competition from other countries and international organizations. In fact, one of the reasons Japanese ODA evaluations received high marks in the 1996 OECD-DAC (Development Assistance Committee) Peer Review on Japanese ODA is that Japan has a relatively long history of evaluating ODA activities.

In the process of constructing the Japanese government policy evaluation system, the practice of evaluating ODA, which had already been established to a degree, was considered as an area where evaluation should be mandatory. The 1998 “Memorandum of Understanding by the Managers of Ministerial Meetings on International Economic Cooperation” stipulated that ministries and agencies involved in ODA should, from the perspective of improving transparency and efficiency: (i) improve evaluation systems and promote information disclosure; (ii) properly implement ex-ante studies and various evaluations; (iii) strengthen monitoring at the stages of ex-post evaluation and implementation; and (iv) rigorously utilize evaluation results in projects and programs (Institute of Administrative Management 2006). Now, under the Government Policy Evaluations Act, ex-ante evaluations are required for projects and programs costing more than a certain amount.

Meanwhile, the national budget shortfall and the push towards comprehensive government reform in recent years have led to demands for a transition from quantitative growth of ODA to qualitative improvement, making evaluation even more important for making aid more efficient and effective. The new ODA Charter (MOFA 2003), established in 2003, regards “enhancement of evaluation” as essential, noting:

“The government will carry out consecutive evaluations at all stages, i.e. ex-ante, mid-term, and ex-post, and at all levels, i.e. policy, program, and project. Furthermore, in order to measure, analyze, and objectively evaluate the outcome of ODA, third-party evaluations conducted by experts will be enhanced while the government undertakes policy evaluations. The evaluation results will be reflected in subsequent ODA policy-making and efficient and effective implementation.”

In short, evaluation is becoming more important for ensuring accountability and transparency, and as a tool for learning and improvement with the

goal of enhancing the quality of ODA. Furthermore, the move towards enhanced information disclosure requires greater efforts for ensuring transparency. Further disclosure of evaluation results is essential to ensure ODA's transparency and accountability and in turn gain the people's understanding and support for ODA.

This chapter discusses the current status and issues of evaluating ODA in Japan in regards to the government's attempt to enhance evaluations in response to these recent trends for greater accountability and transparency. The next section provides an overview of the history of Japan's ODA evaluation and looks at how evaluation systems and institutions have been established. Section 3 examines basic evaluation policies and the current status of ODA evaluation. Section 4 discusses remaining issues and how they could be resolved. The final section describes the future prospects for ODA evaluations.

2. Japan's ODA Evaluation Systems

2.1 Move towards establishing ODA evaluation systems

(1) Foundation of evaluation systems

With the increasing importance given to ODA evaluation, the Council on ODA Reforms for the 21st Century, an advisory body reporting to the Minister of Foreign Affairs, released its final report in January 1998 which pointed out that "establishing evaluation systems" was important for the purpose of constructing more efficient ODA implementing systems. In response to this finding, the ODA Evaluation Reviewing Panel, an advisory body reporting to MOFA's Director-General of Economic Cooperation Bureau, created the Evaluation Working Group in November 1998 to discuss the problems and challenges of ODA evaluation and prepare concrete recommendations. Following the discussions by the Working Group, the Panel submitted the *Report on Reform of Japan's ODA Evaluation System* to the Minister of Foreign Affairs in March 2000.

This report presented concrete reform proposals based on systematic and comprehensive discussions about ODA evaluation in terms of "for what" (objectives), "what" (subjects), "when" (timing), "who" (responsibilities and human resources), "how" (systems and methods) and "how to utilize" (feedback, public relations). This was the first attempt by a Japanese government agency to seriously discuss the basic concepts of ODA evaluation. Subsequent aid evaluation reforms were shaped by the recommendations of

Table 1. Main recommendations of the Report on Reform of Japan's ODA Evaluation System (March 2000)

① Subjects of the Evaluation	In addition to existing project- and program-level evaluations, policy-level evaluations should be introduced. Also, evaluation on the fields and programs of growing importance, areas for which evaluations have not been fully conducted, should be promoted.
② Responsibilities of the Evaluation	MOFA should focus on policy-level evaluations rather than the evaluations of individual projects. JICA and JBIC should promote the improvement of project-level evaluations.
③ Institution Strengthening in the Evaluation	MOFA, JICA, and JBIC should have evaluation specialists in each evaluation division/unit. These specialists are expected to be in charge of the evaluation for a longer period and understand the entire scope of the ODA evaluation. The scope and responsibilities of the "external (third-party) specialists" should be expanded and active and practical use of think tanks and consultants should be promoted.
④ Human Resources Development in Evaluation	Overseas training and the scholarship programs should be enhanced, and the specialized education system of ODA evaluation should be strengthened in postgraduate courses and research and educational institutes related to international assistance. In addition, practical and effective use of human resources should be upgraded. A registration system for evaluation specialists could also be introduced.
⑤ Timing of Evaluation	A consistent evaluation system, from the ex-ante and mid-term evaluation to the ex-post evaluation, should be established.
⑥ Evaluation Methods	The evaluation method based on the "DAC Evaluation Principle" which uses five evaluation criteria should be upgraded, and the evaluation items and viewpoints need to be enhanced. The analysis method of the socio-economic impact should be strengthened in order to realize effective and efficient project implementation.
⑦ Practical Use of Feedback	Evaluation feedback systems should be further enhanced, and a feedback cooperation system among the aid agencies should be established.
⑧ Information Disclosure and Public Relations on Evaluation	Evaluation report formats should be coordinated and unified as much as possible and input into a database, and broader and timelier access to evaluation results should be promoted by posting them on websites. Opportunities should be extended where citizens, NGOs, local governments, and local assembly members can participate in the evaluation (especially monitoring) activities.

Source: Muta (2004)

this report. Its main recommendations are shown in Table 1 (ODA Evaluation Reviewing Panel 2000).

To discuss further details, the ODA Evaluation Study Group was created under the ODA Evaluation Reviewing Panel. The Group met eight times and discussed five issues: ① Introduction of policy-level evaluation and expansion of program-level evaluation; ② Strengthening of the evaluation feedback system; ③ Development of evaluators and effective deployment; ④ Ensuring consistency in evaluation (establishment of a consistent evaluation system, from the ex-ante and mid-term evaluation to the ex-post evaluation);

and ⑤ Promoting collaboration among ministries involved in ODA. The Study Group's 14 members (including; representatives of the academia, economic organizations; NGOs; and international organizations) had substantial and technical discussions together with observers representing all ODA ministries (17 at the time) and the Board of Audit. Their output was submitted as the Report of the ODA Evaluation Study Group, *Improvement of the ODA Evaluation System*, to the Minister of Foreign Affairs. In their discussions, the two groups covered nearly all major issues of the ODA evaluation system, and built the foundations for the current evaluation system.

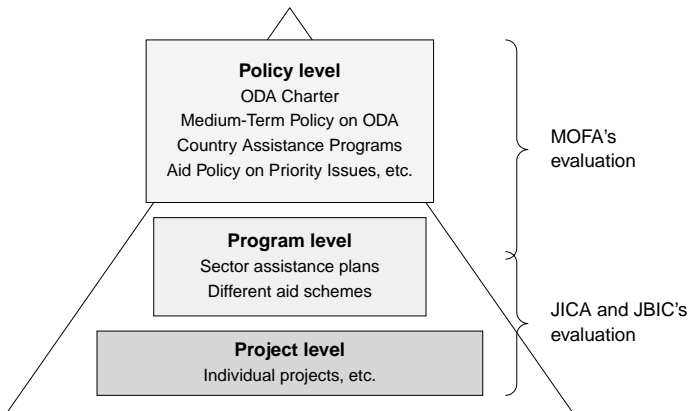
2.2 Organizations that conduct ODA evaluations

(1) Responsibilities for evaluation

In Japan, ODA evaluations are mainly conducted by three organizations: MOFA, JICA, and JBIC. MOFA is responsible for ODA policies whereas JICA and JBIC are implementing agencies.

Since MOFA is responsible for making economic cooperation policies, it is mainly in charge of policy- and program-level evaluations. On the other hand, JICA and JBIC, as implementing agencies, mainly evaluate individual projects. However, because MOFA has administered the bulk of main part of grant aid budgets, MOFA has been conducting ex-post evaluations of grant aid cooperation projects since FY2005. JICA and JBIC also conduct thematic and sector-wide evaluations deemed necessary for strategic reasons by implementing agencies. The organizations that conduct ODA evaluations and

Figure 1. Organizations that conduct ODA evaluations and their evaluation subjects



Source: MOFA International Cooperation Bureau (2007: 10)

their evaluation subjects are shown in Figure 1.

Table 2 shows the numbers of evaluations conducted by each organization in FY 2005.

Table 2. Evaluations conducted by MOFA, JICA, and JBIC (FY2005)

	MOFA	JICA	JBIC
Policy-level evaluation	6	—	—
Program-level evaluation	7	—	—
Project-level evaluation	Ex-post: 52 (grant aid)	Ex-ante: 109 Mid-term: 24 Completion: 73 Ex-post: 46	Ex-ante: 50 Mid-term: 10 Ex-post evaluation: 56 Ex-post monitoring: 7
Thematic evaluation	—	6	4

Sources: MOFA International Cooperation Bureau (2006), Japan International Cooperation Agency (2006) and Japan Bank for International Cooperation (2006)

(2) Reform of ODA implementation organizations and evaluation system

In August 2006, MOFA underwent organizational changes designed to strengthen ODA planning functions. Part of the changes also included integrating the Economic Cooperation Bureau and a part of the Multilateral Cooperation Department into the new International Cooperation Bureau. Changes were also made with regard to the department in charge of comprehensive ODA evaluations that saw the Evaluation Group within the Development Planning Division being upgraded to the independent ODA Evaluation Division.

In addition, in 2007 as part of ODA reform, the Parliament passed a law to amend the JICA Law. In October 2008, JICA and the overseas economic cooperation divisions of JBIC was integrated into the new JICA which now implements and administers; technical cooperation, Yen Loans and most of Japan's grant aid cooperation programs. After the merger, evaluations that used to be implemented by separate implementing agencies are now conducted under a unified mechanism, which should further enhance the evaluation system (MOFA International Cooperation Bureau 2007).

(3) Collaboration among ODA ministries

MOFA, JICA and JBIC together account for slightly more than half of the

government's entire ODA budget. The rest of the budget is distributed to other ministries and agencies which use most of it for human resource development programs such as dispatching experts, research studies, and training courses and seminars. To improve the quality of ODA, evaluation of ODA programs should be considered which span across the entire government. In fact, to establish an ODA evaluation system for the entire country, the Liaison Meeting of Evaluation Divisions of ODA-Related Ministries and Agencies (currently ODA Evaluation Liaison Meeting) was organized in July 2001 as a forum for the regular exchange of opinions and discussion among ODA-related ministries and agencies.

The Liaison Meeting is supposed to consider the preparation of standardized guidelines, manuals, and templates that can be used by all ministries and agencies. Currently, ministries other than MOFA also evaluate their ODA activities and publish reports which makes it difficult even for the officials in charge of ODA-related ministries (not to mention the general public) to know what kinds of evaluations are being conducted by other ministries. In this regard, increased efficiency of ODA as a whole, including evaluation activities, is not possible without the collaboration of MOFA, JICA, JBIC, and all other related ministries.

3. Basic philosophy and present state of Japan's ODA evaluation

As described above, evaluation of Japan's ODA is primarily undertaken by MOFA and the two implementing agencies: JICA and JBIC. This section takes a look at these organizations, reviews the basic philosophy of Japan's ODA evaluation, and examines the present state of ODA evaluation. The basic philosophy of aid evaluation that has become mainstream will be examined with four organizing concepts: ① objectives of evaluation, ② evaluation methods, ③ results-oriented approach, and ④ consistent ex-ante, mid-term and ex-post evaluations.

3.1 Objectives of evaluation

Different organizations define objectives of evaluation differently but definitions can be categorized into two groups as follows:

(1) Ensuring accountability and transparency

Aid activities are financed by taxes and donations. The fundamental objective

of evaluation is to explain clearly to taxpayers and donors how their money was used and what was achieved. Evaluation also ensures transparency. In the 1990s Japan became the world's top donor country in terms of aid volume. This was due in part to increased ODA spending by Japan, but also due to the "aid fatigue" of other developed countries. Aid fatigue, caused in part by a country's doubt over the effectiveness of aid, leads to a stagnation or decrease in a country's ODA spending.

As Japan entered the new century (2000), Japan also saw a decrease in its ODA spending due to lingering doubts about aid effectiveness as well as the prolonged economic slowdown which has made assistance more expensive and difficult to justify. Therefore, in order for Japan to gain the support of its people and continue assistance in a stable manner it has become even more important to answer questions such as "Is our aid making a difference?" and "Is it being implemented efficiently?"

It has become the norm to publish results of ex-ante, mid-term and ex-post evaluations on websites or through other media. Also, to ensure greater objectivity in ODA evaluations which are for the most part conducted internally, MOFA, JICA and JBIC are conducting secondary evaluations based on the original evaluation reports. Conducting secondary evaluations is also an attempt to evaluate their assistance programs as a whole. (Information on these evaluations is also widely published on websites, etc.)

(2) Learning and improvement

Not all aid activities are successful, and unsuccessful activities should be improved through a learning process. If the result of an evaluation indicates that the initial targets have been achieved, continuing the existing way of doing things can be justified. If problems are uncovered, however, they must be fixed. An evaluation of an already completed aid project can be useful for similar subsequent activities. Evaluations can also play a part in operational control of an on-going project and directly contribute to improving the project itself. Therefore, enhanced evaluations will, in the long run, ensure better aid quality.

Currently, organizations in charge of evaluation are engaged in various initiatives. For example, JICA is utilizing past evaluation results in ex-ante evaluations conducted at the project planning stage by requiring its staff to cite evaluations of similar projects in terms of the lessons learned and recommendations made. JICA is also publishing summaries of past evaluation experiences in a website called "JICA Knowledge Site." For its part, JBIC is verify-

ing how the results of ex-post evaluations are utilized in relevant projects through a mechanism called “ex-post monitoring.”

While it is true that evaluation reports in the past tended to be underutilized (JICA Evaluation Department, Office of Evaluation and Post Project Monitoring 2001; Muta 2004), mechanisms through which evaluations actually lead to improvements are beginning to be put into place. For example, broad recommendations based on discussions of evaluation reports including MOFA’s External Advisory Board on ODA Evaluation, JBIC’s Yen Loan Evaluation Expert Committee, and JICA’s External Expert Operations Evaluation Committee are eliciting necessary responses and follow-up activities (MOFA International Cooperation Bureau 2007).

3.2 Evaluation criteria

There are five commonly used points of aid evaluation which are based on the five evaluation criteria published by the Development Assistance Committee (DAC) of the Organization for Economic Cooperation and Development (OECD) in *Principles for Evaluation of Development Assistance* (JICA Planning and Evaluation Department, Office of Evaluation and Post Project Monitoring 2004). They are: ① Relevance (appropriateness and necessity of the aid project), ② Effectiveness (whether the project really has impacts on beneficiaries or the society), ③ Efficiency (whether resources are being utilized effectively, mainly focusing on the cost-effect relationship), ④ Impact (longer-term, indirect effects and ramifications of the project), and ⑤ Sustainability (whether benefits of the project are sustained even after donor activity is terminated).

These five evaluation criteria are widely used in project-level evaluations, although the relative emphases placed on them are different depending on; the timing of evaluation (ex-ante, mid-term, ex-post); and the nature and status of the evaluated project. Conducting evaluations with the common criteria facilitates the accumulation and organization of ODA project evaluation data and allows more efficient use of evaluation results. On the other hand, these five criteria do not necessarily work well in evaluations at a higher (program or policy) level. For this reason, there are many examples of adopting different criteria - deemed appropriate for a specific evaluation - on a case-by-case basis.

3.3 Results-oriented evaluation

(1) Evaluation of outcome and impact

Results of development activities are conceptually categorized into output, outcome and impact depending on the degree of causality and social influence. Both the terms “outcome” and “impact” are used to explain degrees of influence that a concrete result has in a society, and it is difficult to precisely distinguish between them. In many cases, they are used to mean almost the same concept depending on the scope and degree of its influence. In an ODA evaluation, one is required to evaluate the logical process in which various inputs of aid activities produce direct results (outputs), and these outputs function in a society to become outcomes, and then these outcomes produce impacts which represent the ultimate objective of the aid activity.

Let us consider the example of an aid project for building schools (the type of project often implemented by Japan). In this case, the output would be the number of newly built schools; the outcome would be an increase in the enrollment ratio; and the impact would be attainment of equal opportunity for and improved quality of education. If one is to evaluate a school construction project, he/she must, of course, evaluate whether the schools were built as planned but additionally he/she should consider whether or not the construction of schools led to an increase in the enrollment ratio and helped expand educational opportunities, and moreover, whether the new comfortable learning environment actually improved the quality of education. In addition, it would be desirable if enrollment increased faster than the target population and produced a higher enrollment ratio, and if educational opportunities expanded equally in urban and rural areas as well as equally in terms of gender. However, contrary to expectations, there could be problems such as insufficient enrollment in the newly built schools, high dropout and grade-repeat ratios, and stagnant achievement levels. Many examples exist in which the enrollment ratio did not improve or educational targets were not achieved. Reasons may include; too much school capacity relative to the school age population in the target area; inconvenient school locations; inability to hire enough qualified teachers; lack of support for education on the part of local communities and parents; and lack of teaching materials that prevents effective instructions. In such cases, building schools will not accomplish very much because completing the construction of the physical facilities does not necessarily mean that the assistance produced the desired results.

It is now commonly accepted that the true results of assistance cannot be evaluated by evaluating only the outputs and ignoring outcomes and impacts. The concept of results-oriented evaluation emphasizes, more than anything else, the importance of clearly defining the ultimate goals and identifying actual results (CIDA 1999; Japan Institute for Overseas Investment 2001; Kusek and Rist 2004). It is important to have a clear understanding that, for any project, inputs will produce the desired outputs which will be transformed into outcomes and impacts. Before implementing an individual, specific project, the mechanism and logic through which the project produces outputs and creates outcomes must be clearly understood. The Logical Framework Approach and Project Design Matrix (PDM) are used by JICA and JBIC as a tool to design that logic, and at the same time, as an important source of information for choosing the specific evaluation criteria.

(2) Sector-wide evaluation

Outcomes and impacts are higher-level objectives. There are usually various routes (ways) to achieve any given higher-level objectives and many routes may eventually lead to the same outcome. However, some routes may be more difficult than others. In fact, some projects produce desired outputs but fail to achieve expected outcomes. In some cases, it may take a long time before outcomes manifest themselves. Also, in many cases, project design is so poor from the very beginning that it's unlikely that outcomes will ever manifest themselves.

Again, let us consider an education project as an example. The ultimate objectives of any educational assistance project (especially at the basic education level) are quantitative expansion and qualitative improvement. If one considers what can be done for quantitative expansion, i.e., to increase the enrollment ratio and expand educational opportunities, there are many routes in addition to building schools which can be taken to get to the ultimate objective. Examples include; promoting parents' appreciation for education so that they encourage their children to enroll; implementing projects designed to support the enrollment of girls whose enrollment ratio is usually lower than that of boys; providing incentives to come to school such as school lunch programs; providing enrollment assistance to the poor; and starting double sessions in existing schools (instead of building new schools). Given all these alternatives there is no guarantee that building new schools will always be the most effective route. In some situations, a school building project alone may not lead to the achievement of the ultimate objec-

tives at all. Instead of building schools, it may be more effective to train teachers or have enrollment campaigns. Or, it may be best to incorporate teacher training and enrollment campaigns into a school building project. This is why an evaluation that looks at the entire education sector is necessary. There can be no *a priori* conclusion that school construction is the best solution. One can only claim the appropriateness of a school construction project after having comparing it to other projects with similar high-level objectives and determining that it will be the most effective and efficient project for expanding educational opportunities and improving the quality of education.

As it stands, it is becoming increasingly important to evaluate not only individual projects but also programs and sectors under a larger framework and from a higher perspective.

(3) More comprehensive evaluation

If a single project is unlikely to produce sufficient outcomes and impacts, one must, by necessity, make aid more comprehensive by combining different methods. If aid becomes more comprehensive, outcomes and impacts are more likely to be achieved. While comprehensive aid combining various methods sounds good in principle, Japan may not be able to do it all by itself. To really achieve outcomes, it is important to continue encouraging the recipient country to play their role properly.

Thus, results-oriented evaluations have expanded beyond the project level into specific sectors, issues, and overall results of assistance. Efforts to set specific numerical targets for improving development indicators and to achieve the results through comprehensive approaches was spurred by a series of initiatives begun in the late 1990s including the DAC New Development Strategy, the Comprehensive Development Framework (CDF), Poverty Reduction Strategy Papers (PRSPs), and Millennium Development Goals (MDGs) (Miwa 2007).

As mentioned above, although program-level and policy evaluations are conducted every year mainly by MOFA, systematic evaluation methods are not quite as established as those of project-level evaluations.

3.4 Consistent ex-ante, mid-term and ex-post evaluations

An ex-ante evaluation, a mid-term evaluation (mid-term review in JICA) and an evaluation at the time of completion or ex-post evaluation are conducted for basically every ODA project based on the recognition that such evalua-

tions based on consistent criteria are necessary in order to make evaluation activities more effective. An ex-post evaluation cannot be conducted fully unless various indicators and data are prepared before the project begins. The actual performance of a project can only be measured by conducting an ex-ante evaluation using the same criteria as those of an ex-post evaluation and noting the changes in the indicators based on initial measurements.

In May 2001, JICA announced that it would prepare and publish ex-ante evaluation tables with regard to General Grants, Fisheries Grants and Project Type Technical Cooperation projects. JBIC did the same with regard to its Yen Loan projects. In these tables, numerical targets are also described. It has now become common to prepare a Logical Framework or PDM at the start of a project. While these are desirable developments, there is still insufficient analyses which examines whether the planned project has an advantage over other possible projects that could also achieve the same outcomes or higher-level objectives. We must think about obtaining necessary data through preliminary studies, understanding the linkages between factors that contribute to the desired results, and designing a project that adequately incorporates those important factors. That means the success of an ex-post evaluation depends on how much time is spent before the beginning of the project to increase its “evaluability” (CIDA 2000).

A mid-term evaluation or review is conducted to verify whether an ongoing project is being implemented as planned and to determine whether there are any potential factors that may prevent the achievement of expected results and if necessary make adjustments. Although it is of course necessary to try to achieve numerical targets established by the ex-ante evaluation, in general, the process from planning to project completion takes many years, and conditions change. Even the most carefully designed project may not proceed as planned. To respond to such situations, the mechanisms of mid-term evaluation and monitoring are used to incorporate evaluation results appropriately and make necessary adjustments. Currently, most mid-term evaluations are conducted by the internal staff, but there is a need to consider mid-term evaluations conducted by evaluators including external, third parties, and establish official procedures to change target values in order to maintain transparency while administering projects effectively under realistic conditions.

JBIC conducts “ex-post evaluations” on all implemented projects two years after the completion date. These evaluations examine factors such as the efficiency of the implementation method and the relevance and sustain-

ability of the project. The results are fed back to the recipient country's implementing agencies. JICA conducts an "evaluation at the time of completion" on every project immediately before it is completed. Also, in FY2002, JICA introduced the "ex-post evaluation" which evaluates a project after a stated amount of time has passed since the termination of cooperation. It mainly verifies whether the impacts of cooperation have been sustained and whether long-term and/or indirect impacts have manifested.

As it stands, both implementing agencies have introduced ex-ante, mid-term and ex-post evaluations based on consistent criteria.

4. Remaining issues of Japan's ODA evaluation

As described above, the ODA Evaluation Reviewing Panel was established in 1998 as an advisory body for MOFA's Director-General of Economic Cooperation Bureau (at the time). The Evaluation Working Group and the ODA Evaluation Study Group, established under the Panel in 1998 and 2000, respectively, have sorted out main issues surrounding the evaluation of ODA, and attempts have been made to establish evaluation methods and systems. These discussions are reflected in the basic philosophy of ODA evaluation we summarized in the previous section. Also, in recent years the Investigative Commission for the Evaluation of Medium-Term Policy on ODA (2004) and other groups have recognized the need for program and policy-level evaluations and have been discussing how higher-than-project-level evaluations should be conducted.

Based on recent discussions, this section discusses the remaining challenges of ODA evaluation in terms of five issues: ① evaluation methods, ② subjects of evaluation, ③ evaluation feedback, ④ collaboration among ODA-related agencies, and ⑤ development of evaluators.

4.1 Improvement and development of evaluation methods

(1) Clarification of targets and evaluation indicators

An evaluation is basically a comparison of results with the original plan, and they are two sides of the same coin. In evaluating ODA, the Logical Framework and PDM are used and results-oriented evaluations are conducted with the premise that targets have been defined as clearly as possible at the planning stage. The point of evaluation is not to find out whether those involved in the project did their best, but to measure and demonstrate to

what extent those targets, in terms of indicators, have been achieved. Whenever possible, one needs to quantify not only expected outputs but also outcomes and impacts. If quantification is impossible, then outputs, outcomes and impacts should be specifically described. Although target values should be clearly defined before a project begins, this is not always the case.

Behind this shortcoming is a concern that setting indicators may unduly limit the scope of project activities. Some also point out that setting numerical targets at the planning stage of “soft” projects such as capacity building projects is difficult. However, in many cases, the real difficulty is caused by the lack of clear agreement among those involved to determine what the project’s concrete action targets are and what specifically they are trying to change by the project.

What is important is that aid professionals establish a process to adequately discuss and agree on the targets of each project, whether those targets are quantitative or qualitative. Evaluation results are more convincing if they offer quantitative analyses that can be concretely and objectively explained. Even with regard to “soft” projects in social development and other supposedly difficult to quantify areas, it is indeed possible to quantify their qualitative aspects. Qualitative evaluation is also a viable alternative. While unreasonable quantification is unnecessary, efforts for quantification help make issues clear. Utilizing a healthy balance of both quantitative and qualitative data in an evaluation contributes to learning and improvement which are fundamental objectives of evaluation. More efforts are needed to clarify numerical targets and quantify qualitative aspects of evaluated projects, or utilize qualitative evaluation techniques.

(2) Evaluation of efficiency and cost

Evaluation reports have been weak on cost and efficiency analyses and this aspect needs to be strengthened. Judgment of efficiency is basically a matter of comparison. In reality, an individual project is not being examined adequately in terms of whether its interventions were better than the interventions of other projects, let alone whether the project’s interventions were the best way to achieve the targets. As an increase in ODA budget is becoming more unlikely and demands for better results are getting stronger, efficiency is becoming more important. There is an urgent need for organizing data that can be used as a reference for comparison.

If large amounts of inputs are thrown into a situation where resources are limited, as in most aid recipient countries, of course some results can be

expected. However, if excessive amounts of inputs are needed to obtain results, the cooperation is not sustainable, because the recipient country will not be able to ensure the same amount of inputs by itself (Muta 2003). In other words, we should keep in mind that even if a project is cost-effective in terms of cost-benefit analysis, it could be problematic in terms of sustainability if it requires huge amounts of costly inputs.

(3) An attempt to establish a rating system

In FY2004, JBIC launched a full-fledged rating system designed to produce quantitative evaluation results. The system assigned a rating of A through D to each of the five evaluation criteria. In FY2006, the effectiveness of the system itself was examined based on the past rating results, and a new system with more detailed, 25 criteria, was introduced on a pilot basis.

The attempt to break down the five evaluation criteria shows what kinds of conditions are necessary for a project's success based on the characteristics of Yen Loans. This attempt is interesting and important not only for the clarification of lessons and recommendations through evaluation but also for project preparation and implementation. There is hope that this sort of development and trial of new evaluation methods based on the analyses of past evaluation experiences will lead to more objective and accurate evaluations and more fine-tuned lessons and recommendations.

4.2 Dealing with new evaluation subjects

(1) Program- and policy-level evaluation

At the beginning of this chapter we described the long history of ODA evaluation. Most of this history involves the evaluations of individual projects designed to, for example, build infrastructure such as roads and dams, build facilities such as schools, and transfer agricultural development technologies. In recent years, however, many people began to question whether these project-level evaluations were sufficient in light of the advancement of country and thematic programs designed for more effective aid as well as the increased attention to results in certain thematic areas and country-level development (Muta 2004; Miwa 2007). The necessity for higher-level evaluations is now widely recognized for; program-level evaluations which comprehensively examine multiple projects that belong to the same sector or have common thematic objectives (such as poverty, gender, primary education, structural adjustment loans); and evaluations performed at an even higher

level than the program level, which examine various aid policies of Japan (including the Medium-Term Policy on ODA and Country Assistance Programs).

Conventional, project-oriented evaluation methods are inadequate for program- and policy-level evaluations. There are some practical problems. First, even if we understand conceptually the importance of program- and policy-level evaluations, it is not easy to draw clear lines between them. Secondly, while the necessity of program- and policy-level evaluations is internationally recognized, and while some foreign organizations have made attempts to conduct them, unified, concrete methods have yet to be established. What is needed now is an effort to examine past practices of ODA evaluation and develop and enhance evaluation methods that are suitable for Japan's ODA.

First, the policy-level evaluation assumes that aid activities are carried out with clear objectives at the policy level. If multiple projects are carried out only with project-level objectives, bundling them up and evaluating them would not be, strictly speaking, a program- or policy-level evaluation, because the plurality of projects does not constitute a coherent structure of a program, and because the plurality of programs as a whole does not reflect a policy. In its recommendations, MOFA's ODA Expert Council notes that Country Assistance Programs need to be clarified, and logical structure to clarify the effects of aid should be given greater emphasis (MOFA International Cooperation Bureau 2006).

The same can be said with regard to the program-level evaluation. Before sector-wide evaluations can be performed, systematic goals must be prepared for each sector, and assistance programs must be designed in line with these goals. In other words, policy- and program-level evaluations require that, for each of Japan's aid policies, Country Assistance Programs, programs and projects, goals and targets are narrowed down and indicators are clearly established at the earliest possible stage. To this end, it is effective to introduce a target system schematic at the planning stage, and it is essential to establish and regularly monitor targets and corresponding evaluation indicators at each level.

Although some Country Assistance Programs prepared since 2003 include policy diagrams, the numbers are still inadequate. From now on, we need to plan projects in line with assistance programs and sector programs which are, in turn, based on such structured target systems. When planning projects, it is also necessary to adequately consider the coordination with other aid organizations and the consistency with the developing country's

own development strategy, because at a higher level of targets, factors, other than Japanese aid agencies' activities, are more influential, and collaboration with related organizations are more important.

(2) Expansion of projects to be evaluated

There are still quite a few areas and types of ODA projects that are not adequately evaluated. For example, evaluations are not adequately performed on; training and scholarship programs; the dispatch of experts; the JOCV (Japan Overseas Cooperation Volunteers) programs; Grassroots Grants; contributions to international organizations; and the assistance of NGOs. Reasons for this include the fact that these projects deal directly with people and are therefore difficult to evaluate, and that the budget for each project is so small that it is difficult to justify spending money on evaluation. However, since these projects as a whole constitute an important part of ODA, we must think about expanding the scope of evaluation to cover these areas.

Success of a project depends largely on human factors. While there are difficulties associated with an expert going to a foreign country and carrying out his assignments, the expert him/herself must be evaluated as an important component of the project. In an evaluation of an expert, one should not focus on judging the abilities of the individual. Instead, it is more important to evaluate whether his/her abilities matched the needs on the ground, and to provide feedback on the expert recruiting strategies and policies in terms of how to recruit experts who have the required abilities, and what kinds of experts are suitable for a certain geographical area, etc.

Programs directly related to human development such as training, student-exchange and cultural exchange programs have, in the past, tended not to be evaluated because their results can only be seen in the long run. However, in these difficult fiscal times, and considering that a large amount of money is being spent for these programs as a whole, it is becoming increasingly difficult to explain that we must wait 10, 20 years before we can see their results. While some results may take 10 years to manifest themselves, some results must begin to show their "buds" in one or two years. Even if it is impossible to evaluate 100% of the results, it is important to try to find the "buds" of results on prehensible matters and evaluate them, even if they represent only 5% or 10% of the entire results. Even in the area of human development, there are many projects that have been completed more than 10 years ago. The fact that it takes time to see the results is insufficient reason not to evaluate them.

4.3 Enhancing the evaluation feedback system

The work of evaluation itself is finished with the completion of an evaluation report. However, the lessons and recommendations written in that report will not automatically be utilized. It is quite common to see evaluation reports remain unread. Even when they are read, it is rare to see them put to good use and their recommendations reflected in concrete activities. On the topic of how useful evaluations have been for increasing the efficiency and effectiveness of aid, some problems have been pointed out, including insufficient PR activities to promote the evaluation results and unclear positioning of evaluations in the project cycle. The Swedish International Development Agency (SIDA), for example, stipulates that evaluation results should be considered and reflected in the policy-making process as well as in new and ongoing aid projects (SIDA 1999), but in reality, results are rarely utilized (Carlson 1999). The analysis shows that in some cases evaluation results are not even communicated to the recipient country. Even when they are utilized, they are not directly used for improvement; instead, they are simply used for understanding relevant concepts and justifying aid activities. In the 2001 survey of JICA employees and experts on the use of ex-post evaluations, many respondents answered that they did not utilize the results of ex-post evaluations. The reasons given for not utilizing results included; “not aware of ex-post evaluations themselves,” “not knowing how to obtain them,” and “work can be done without using them.” These reasons reflect both the problem of insufficient PR promotion of ex-post evaluations as well as the problem of unclear positioning of evaluation in JICA’s project cycle (JICA Evaluation Department, Office of Evaluation and Post Project Monitoring 2001).

While it is important to promote the awareness of evaluation data and to improve their quality to make them easier to use, these efforts alone will not lead to learning and improvement which are the fundamental objectives of evaluation. It is important to create and strengthen mechanisms within organizations and an ODA implementing systems that can adequately utilize and incorporate the feedback in operations. To that end, evaluation divisions must cooperate with planning and operations divisions to promote the feedback. For example, in 2003 MOFA reorganized the existing Evaluation Feedback Committee and created the External Advisory Board on ODA Evaluation. Since then, all evaluations by MOFA are implemented by the Board, and its recommendations are given to the Economic Cooperation Bureau (now called the International Cooperation Bureau). Within the bureau, an internal review committee creates action plans in response to the

recommendations and submits these plans to the Advisory Board for its approval. Furthermore, the statuses of implementation of these action plans are published in the annual evaluation report. It is necessary to create a permanent “evaluation feedback committee” within each organization which includes the participation of executives and members of the planning, operations and evaluation divisions and which monitors the evaluation results and how the results are utilized. To make good use of the feedback, it’s essential to construct an appropriate mechanism and then strictly administrate it.

Also, qualitative improvement of ODA requires feedbacks for ODA as a whole. In other words, it is important to establish a collaboration system for sharing feedbacks among all ODA implementing agencies in Japan. We can hope that the interministerial coordination body, the ODA Evaluation Liaison Meeting will be the basis for such a system. It would also be a good idea to construct a database that centrally controls the data from the results described in the evaluation reports of MOFA, JICA, JBIC and other ODA-related ministries, and to create a system for sharing the database.

Furthermore, we must not forget about providing feedbacks to aid recipient countries. It is important to make absolutely sure that evaluation results are officially communicated to the aid recipient country, and to support the country’s efforts to incorporate the lessons and recommendations in the preparation and implementation of future projects and programs. In 2000, a workshop of OECD-DAC Working Party on Aid Evaluation was held in Tokyo, and for the first time, observers from developing countries were invited. Also, at the ODA Evaluation Seminar held at the same time, the importance of cooperation on evaluation between donors and aid recipient countries was emphasized (MOFA Economic Cooperation Bureau, ODA Evaluation Division 2000). In the past, ODA evaluations tended to be conducted by donors. The participatory evaluation, which involves stakeholders in the recipient country in evaluation activities, will enhance the evaluation capabilities of the recipient country and also contribute greatly to feeding the evaluation results back to the frontline of development.

4.4 Development of evaluators

When the importance of aid evaluation is recognized and the field is energized, human resource development becomes an important factor for improving the quality of evaluation. Unless conducted by experts who have a certain level of specialized knowledge and skills, an evaluation may end up as a simple critique, and the results will not be trustworthy. For example, JICA’s sec-

ondary evaluation showed that the quality of evaluations was poor because of insufficient data gathering through surveys and interviews (JICA 2006). The same evaluation also noted the lack of objectivity and logic in evaluation reports. The lack of objectivity being due to the lack of clear explanation caused by poor writing and insufficient explanation of the evaluation's quantitative and qualitative analyses.

To improve the quality of evaluations and the authority of their results, expert evaluators are necessary. There is an urgent need to develop evaluators by; creating long-term courses and enhanced professional education programs in graduate schools; developing and expanding short-term training programs for practitioners; and offering overseas training and scholarship programs to employees of aid implementing agencies, external experts and consultants. It may be a good idea to certify various evaluation training programs and their trainees in an effort to popularize evaluation and maintain quality. Recently, senior staff members are increasingly allowed to have junior staff members such as graduate school students accompany them in aid evaluation activities, which effectively functions as a kind of on-the-job training.

In September of 2000, The Japan Evaluation Society was established with a purpose of improving the expertise and qualifications of evaluation specialists. In FY2007, the Society launched the Certified Professional Evaluators Training Program which aims to develop highly qualified evaluation specialists. It is hoped that these activities also promote the training and development of evaluators.

4.5 Disclosure and publication of evaluation data

Currently, evaluation reports are published by MOFA, JICA, JBIC, and other ODA-related ministries in different formats. It is important to coordinate and unify the formats as much as possible, and input the reports' information into a database. Also, broader and timelier access to evaluation results should be promoted by posting them on websites. JICA is already publishing its reports in their entirety on its website. Also, when outside evaluators and aid implementing divisions have different opinions on the results of a third-party evaluation, it has become an established practice to include both opinions separately instead of rewriting the report and this has contributed to greater transparency.

People's understanding of and participation in ODA are important so that ODA activities can continue and expand. We should strengthen both our

efforts to expand the opportunities for citizens, NGOs, municipalities and members of local assemblies to participate in evaluation activities and our efforts to build a mechanism through which the general public can freely express their opinions on published evaluation reports and which ensures that these opinions are then reflected in subsequent ODA activities.

Furthermore, it would be a good idea to utilize these evaluation results in the field of education. For example, in Japan, there is a “period for integrated learning” from grade school to high school. Each school can design its own curriculum for this period relatively freely, and children can learn independently. In such classes, evaluators can promote young children’s understanding of foreign assistance by presenting their own evaluation activities. Such activities would be very effective in the long run.

The issues of ODA evaluation described above are summarized in Table 3 below.

5. Concluding remarks and an outlook for the future

In FY2008, with the merger of JICA and JBIC’s overseas economic cooperation divisions, Japan’s ODA implementation system underwent a significant transition. There is a strong hope that integrating the implementing bodies of technical cooperation, grant aid and Yen Loans will lead to more efficient and effective implementation of aid projects. Needless to say, this transition will usher in a new era for ODA evaluation as well. This final section takes a look at the post-merger changes in regards to the evaluation of ODA.

First of all, the merger will allow organic coordination among technical cooperation, loan assistance and grant aid. It is likely that projects that used to be compartmentalized will be planned more coherently with the partner country’s development agenda as the shared, ultimate goal, and as a result, we will move closer to the realization of a true program approach. This will naturally lead to better development results, but it should also raise expectation that program- and policy-level strategies and targets will be defined more clearly, and that structured evaluations based on policy diagrams, showing the routes to achieving targets, will become possible.

Secondly, it is hoped that such structured evaluations will lead to the development of new aid evaluation methods. That is, the enhancement of the program approach through organic coordination among the different aid schemes might enable evaluations that effectively combine; the examination

Table 3. Main issues of Japan's ODA evaluation

1. Improvement and development of evaluation methods	
① Clarification of targets and evaluation indicators	Target values are not necessarily clearly defined. Whether quantitative or qualitative, targets need to be adequately discussed by those involved.
② Enhancing the evaluation of efficiency and cost	To use limited ODA budgets efficiently, the importance of efficiency evaluation is growing. There is an urgent need for gathering data that can serve as a reference for comparison.
③ An attempt at rating	The attempt for rating by JBIC showing the "success factors" of projects more clearly is noteworthy as an effort to develop new evaluation methods.
2. Dealing with new evaluation subjects	
① Enhancing program- and policy-level evaluations	There is an urgent need to establish program- and policy-level evaluation methods. To do so, we need to fully consider the preparation of target system schematics in the policy making process, consistency with the development strategies of developing countries themselves, and collaboration with other aid agencies.
② Expansion of projects to be evaluated	There are still quite a few areas and types of ODA projects that are not adequately evaluated. Examples include training and scholarship programs, dispatch of experts, the JOCV programs, Grassroots Grants and other assistance for NGOs. We need to expand the scope of evaluation to include these areas.
3. Enhancing the evaluation feedback system	
While there has been progress in the publication of evaluation data, there is a need for stronger institutions to connect them to learning and improvement. Also important is a collaboration system for sharing feedbacks not only within an organization but among all ODA-related organizations, as well as providing feedbacks to aid recipient countries. To that end, assistance to improve the evaluation capabilities of recipient countries will be necessary.	
4. Development of evaluators	
To improve the quality of evaluation, highly specialized evaluators are essential. In addition to professional education in graduate schools, evaluation training programs for practitioners should be expanded.	
5. Disclosure and publication of evaluation data	
People's understanding of and participation in ODA are important so that ODA activities can continue and hopefully expand. We need to build and strengthen a mechanism through which feedbacks from the people are actively sought and incorporated.	

of the achievement of impact-level targets; with the examination (or monitoring) of implementation processes and direct outcomes of individual projects. Since evaluations themselves incur costs, there is room to consider alternatives. For example, instead of the more costly alternative of evaluating all projects, the management of individual projects can be enhanced by monitoring and by emphasizing the evaluation of higher-level impacts. The evaluation of impacts is, in a sense, an evaluation of Japan's ODA strategy for a specific development issue, and the results are an important input to the preparation of new strategies. This is clearly different from evaluations of individual pro-

jects conducted with a purpose of supporting the project management. This kind of evaluation provides feedback at the higher, strategy and policy levels, and is conducted with a sector-wide perspective and an awareness of the need for consistency with the partner country's development plans and an awareness of the relationships with other aid organizations.

Aid evaluations examine programs and projects which are essentially outside interventions with a country's development plans. Therefore they require many perspectives that are different from those of domestic, public works evaluations. It is not easy to evaluate how an intervention with the partner country's development process contributed to that country's sustainable achievement of development results. However, by analyzing the evaluation data accumulated through the long evaluation experiences of JICA and JBIC, we should be able to design impact evaluation methods with an understanding of factors that make aid successful and create a set of comparable data for cost-benefit analyses. One example of such analyses is the revised rating system of JBIC.

Also, since evaluation information contains numerous lessons and recommendations as well as valuable data on particular sectors of partner countries, it is hoped that infrastructure for the effective use of such information will be built. After the merger of JICA and JBIC, the new organization will be the sole aid implementing agency in Japan. It may build a database that centrally manages the information written in evaluation reports of not only JICA and JBIC but also MOFA and other ODA-related ministries, and create a system of sharing such data. Some evaluation reports are useful, and some are not. To make them more user-friendly, they should be indexed with easy-to-understand keywords and inputted into a database. Not only will this kind of mechanism lead to more effective feedbacks within aid organizations, but it will also assist in terms of accountability, ensuring transparency, and better public relations. Furthermore, such an evaluation database will serve as a valuable source of information and teaching materials for development education and contribute to the development of human resources for international cooperation. We can expect that these efforts will lead to wider participation of the people, which is necessary for expanding and continuing high quality ODA programs.

As we have seen so far, Japan's ODA evaluation has a longer history and more accumulated experiences than those of domestic policy evaluation. MOFA, JICA and JBIC and their expert panels have discussed and tackled various issues and challenges. Also, beginning in FY2008, Japan's ODA eval-

uation will undergo changes in a new environment. By their nature, ODA projects are implemented under diverse, uncertain and difficult conditions. There is no perfect, impeccable project. What is important is not to talk too loudly about their imperfections, but to determine how evaluation results have improved the evaluated project or subsequent, similar projects. If we don't focus on the effective use of evaluation results to improve ODA projects and programs, ODA evaluation will soon become a mere formality. ODA evaluation must, first and foremost, contribute to the qualitative improvement of ODA. Evaluation is useful only when it promotes social learning by influencing those involved (Picciotto 2000). Now, more than ever, the improvement of ODA will take the efforts of aid professionals, improved awareness, and systematic initiatives to enhance the function of learning through evaluation.

(*This article is translated from “Nihon no kaihatsu enjo hyoka ni okeru kadai to tenbo,” *Kaihatsu enjo no hyoka to sono kadai*, Kaihatsu enjo doko series, 2008, FASID.)

References

Japanese

- Institute of Administrative Management ed. (2006) *Seisaku hyoka handbook – Hyoka shinjidai no torai*, [Policy Evaluation Handbook - A New Era for Evaluation], Gyosei Corporation
- Investigative Commission for the Evaluation of Medium-Term Policy on ODA (2004) *ODA chuki seisaku hyoka*, [Evaluation of the Medium-Term Policy on ODA], Ministry of Foreign Affairs of Japan
- Japan Bank for International Cooperation (2006) *Enshakkan jigyo hyoka hokokusho*, [Ex-Post Evaluation Report on ODA Loan Projects]
- Japan Institute for Overseas Investment (2001) *Segin kokusai kyoryoku ni kansuru hyoka forum hokokusho*, [Report of The World Bank Evaluation Forum on International Cooperation]
- JICA (2006) *Jigyo hyoka nenji hokokusho*, [Annual Evaluation Report]
- JICA Evaluation Department, Office of Evaluation and Post Project Monitoring (2001) *Hyoka kekka no feedback*, [Feedback of Evaluation Results]
- JICA Planning and Evaluation Department, Office of Evaluation and Post Project Monitoring ed. (2004) *Project hyoka no jissenteki shuho*, [Practical Methods of Project Evaluation], Japan International Cooperation Publishing

- Ministry of Foreign Affairs of Japan (2003) *ODA taiko*, [ODA Charter]
- Ministry of Foreign Affairs of Japan, Economic Cooperation Bureau (1999) *Keizai kyoryoku hokokusho (sohen)*, [Economic Cooperation Report (Outline)]
- (2002) *Keizai kyoryoku hyoka hokokusho 2001 nen* [Annual Evaluation Report on Japan's Economic Cooperation 2001]
- Ministry of Foreign Affairs of Japan, Economic Cooperation Bureau, ODA Evaluation Division (2000) *ODA hyoka seminar: Yoriyoi ODA hyoka ni mukete*, [ODA Evaluation Seminar: For Better ODA Evaluations]
- Ministry of Foreign Affairs of Japan, International Cooperation Bureau (2006) *Keizai kyoryoku hyoka hokokusho*, [Annual Evaluation Report on Japan's Economic Cooperation]
- (2007) *Keizai kyoryoku hyoka hokokusho*, [Annual Evaluation Report on Japan's Economic Cooperation]
- Miwa, Satoko (2007) “Kaihatsu enjo hyoka,” [Development Assistance Evaluation], Miyoshi, Koichi ed., *Hyokaron o manabu hito no tameni*, [For Those Who Study Evaluation], Sekai Shisousha, pp.262-282.
- Muta, Hiromitsu (2003) “Kozoteki hyoka ni motodoku sogoteki kokusai kyoryoku no kokoromi,” [An Attempt for Comprehensive International Cooperation Based on Structured Evaluation], *Nihon hyoka kenkyu* 3(1), [The Japanese Journal of Evaluation Studies 3(1)], pp.65-75.
- Muta, Hiromitsu (2004) “Enjo hyoka,” [Aid Evaluation], Goto, Ohno, Watanabe eds., *Nihon no kokusai kaihatsu kyoryoku*, [International Development Cooperation of Japan], Nippon Hyoronsha, pp.137-156
- ODA Evaluation Reviewing Panel, Evaluation Working Group (2000) “*ODA hyoka taisei*” *no kaizen ni kansuru hokokusho*, [Report on Reform of Japan's ODA Evaluation System]
- ODA Evaluation Reviewing Panel, ODA Evaluation Study Group (2001) *Wagakuni no ODA hyoka taisei no kakuju ni mukete*, [Improvement of the ODA Evaluation System]

English

- Carlson, J. et. al. (1999) *Are Evaluation Useful? - Cases From Swedish Development Co-operation*, Swedish International Development Cooperation Agency, Department for Evaluation and Audit, pp.6-8.
- CIDA (1999) *Result-Based Management in CIDA: An Introductory Guide to the Concepts and Principles*, CIDA, Results-Based Management Division, Performance Review Branch.

- (2000) *CIDA Evaluation Guide*, CIDA Performance Review Branch, pp.31-34.
- Kusek, J.Z. and Rist R.C. (2004) *Ten Steps to a Results-Based Monitoring and Evaluation System*, The World Bank.
- Picciotto, Roberto (2000), Concluding remarks, in Feinstein, O. and Picciotto, R eds., *Evaluation and Poverty Reduction: Proceedings from a World Bank Conference*, The World Bank, pp.355-361.
- SIDA (1999) *SIDA's Evaluation Policy*, Swedish International Development cooperation Agency, Department for Evaluation and Audit.

2

ODA Evaluation and Policy Evaluation: Status of Accountability and Transparency in Japan

Kiyoshi Yamaya

1. Introduction

Before examining the respective roles and interrelationship between policy evaluation and the evaluation of Official Development Assistance (hereafter referred to as “ODA evaluation”), we should discuss how “evaluation” is positioned in our country. Precisely because this discussion has not taken place, people often do not realize that they are arguing on different planes. As the saying goes, sleeping in the same bed but having different dreams. For example, when the Cabinet Office pushed for “more efficient public administration” and required objective evaluations for that purpose, officials at the Ministry of Foreign Affairs (MOFA) countered with “diplomatic considerations.” Such mismatches are not uncommon. In fact, there is an overabundance of concepts that use the term “evaluation” in their titles but aim to achieve different purposes. Such concepts include “policy evaluation,” “administrative evaluation,” “administrative project evaluation,” “performance evaluation” and “evaluation of incorporated administrative agencies.” On the other hand, evaluations, not confined to the small circle of public administration such as “ODA evaluation,” “university evaluation” and “school evaluation” are now required to create value, public-private partnerships, and networks in their respective specialized areas. However, evaluations are often conducted without an understanding of who evaluates and for what purpose. This causes confusion, and that confusion is affecting not only policy evaluations but also evaluations in specialized areas such as ODA.

A significant cause of the confusion is that, for the most part, the subjects of evaluation are not clearly defined. Although evaluations may be divided

into subjects such as “policy,” “program” and “project,” there are no clear guidelines for distinguishing one subject from another. The confusion is more serious in the field of domestic public administration than in the field of ODA because, in general, the field of ODA clearly classifies a “policy” as the purpose and direction of activities of the central government, local governments or ministries, and a “project” as a means for implementing the policy. On top of that, a “software” is needed to choose one policy instrument from a pool of many, consider when and how to use it, explain it to those who implement the project, and guide project activities. That software is the “program.” If there is something wrong with this software, the problem that the policy was designed to solve will not be solved. Also, if the program has any problems or bugs, issues will remain unsolved.

The uniquely Japanese terminology, “administrative evaluation” adds to the confusion. Perhaps it is called “administrative evaluation” because the government conducts it, and because it evaluates administrative activities. However, in reality it is administrative project evaluation. As if to further conceal that reality and confuse the discussion even more, the concept of New Public Management (NPM) is often brought up. NPM, however, uses “measurements,” not evaluations. At any rate, when using the vague concept of “administrative evaluation,” further influenced by NPM, policy and program evaluations cannot be distinguished from the performance measurements of administrative operations. The phrase is also difficult to translate into other languages and this makes explaining Japan’s evaluation systems to other countries problematic. It is a very annoying phrase and the situation needs to be clarified.

By organizing the various concepts of evaluation as used in Japan and then by using existing theories of and our experience in ODA evaluation, this chapter attempts to end the confusion by drawing a clear demarcation line between policy evaluation and ODA evaluation.

2. Era of a deluge of evaluations

Japan is now inundated with evaluations. Policy evaluation was first introduced simultaneously in both the national government (the Ministry of International Trade and Industry) and a municipality (Mie Prefecture) in 1997. Since then many trials and errors have followed. Now, the Cabinet, the National Diet, and the ruling party request so many different kinds of “evaluations” and from so many different perspectives that some cynics lament that

Figure 1. Various Evaluation Issues



we are in “an era of a deluge of evaluations.” Evaluation poses a variety of new challenges to every ministry and agency (see Figure 1), and naturally, evaluators on the ground are confused. Moreover, to further confuse evaluators, government auditors have added “effectiveness” to their existing list of evaluation missions which already included traditional criteria such as legality, compliance, economy, and efficiency.

Some evaluators on the ground or in policy-making divisions say that the Administrative Evaluation Bureau of the Ministry of Internal Affairs and Communications (MIC), which is responsible for the system of evaluation, is, in part, the cause of the confusion. They claim, for example, that MIC’s Administrative Evaluation Bureau sometimes requests a ministry to perform multiple evaluations, and repeatedly asks “questions” on the evaluation results submitted in response to the requests. (Table 1 illustrates a past example of MOFA.) In terms of the organizational structure of MIC, different divisions are involved in evaluation: the division in charge of policy evaluation (evaluation to secure coherence/comprehensiveness of policies and evaluation to secure objectivity of policies); the division in charge of administrative evaluation and inspection (which used to be called “administrative inspection”); and the division in charge of evaluating incorporated administrative agencies (IAAs). On the other hand, the General Affairs Division of the Minister’s Secretariat receives these requests and serves as the point of

contact, passing the questions from MIC to relevant divisions through the General Affairs Divisions and Policy Divisions of different Bureaus. The divisions consider the evaluations as “unnecessary additional work, and demand (wrongfully) the General Affairs Division of the Minister’s Secretariat to explain on what grounds they are required to perform such tedious tasks. Although the General Affairs Division of the Minister’s Secretariat may not have a clear understanding of the underlying evaluation requests, they must try to persuade the various divisions to undertake the tasks by citing the relevant laws and regulations which include:

- Policy evaluation is based on Article 4, Section 6 of the Basic Law for Central Government Reform, Article 2, Section 2 of the National Government Organization Law, Article 5, Section 2 of the Act for Establishment of the Cabinet Office, and the Government Policy Evaluations Act (GPEA).
- MIC’s policy evaluation (“evaluation to secure coherence/comprehensiveness” and “evaluation to secure objectivity”) is based on Article 4, Section 17 of the Act for Establishment of the Ministry of Internal Affairs and Communications.
- Public works evaluation is based on Article 17, Section 2 of the Basic Law for Central Government Reform.
- MIC’s administrative evaluation and inspection are based on Article 4, Sections 18 and 19 of the Act for Establishment of the Ministry of Internal Affairs and Communications.
- Ex-ante evaluation of ODA projects, R&D projects, and public works is based on Article 9 of the GPEA and Article 3 of the Order for Enforcement of the GPEA.
- Evaluation of IAAs such as Japan International Cooperation Agency (JICA) is based on the Act of General Rules for Incorporated Administrative Agency and Article 4, Section 19 of the Act for Establishment of the Ministry of Internal Affairs and Communications (including authorized corporations).
- The authority of the Council on Economic and Fiscal Policy which recommends performing various extra works such as systematizing policies and incorporating the results of policy evaluations in budgets is based on Article 18 of the Act for Establishment of the Cabinet Office.

Japan’s evaluation system has expanded through such complicated interorganizational dynamics surrounding the various “evaluations,” but without a full understanding of the purposes and methods of evaluation.

Table1. Evaluations requested by the Ministry of Internal Affairs and Communications to the Ministry of Foreign Affairs (FY2002-04)

		Evaluation subjects	MOFA divisions
Policy evaluation	Evaluation to secure coherence/comprehensiveness	Economic cooperation (2002-04)	Director for Policy Evaluation and Administrative Review, General Affairs Division of Minister's Secretariat, Policy Division and other relevant divisions of Economic Cooperation Bureau
	Evaluation to secure coherence/comprehensiveness	International disaster relief assistance (2004)	General Affairs Division of Minister's Secretariat, Policy Division and International Emergency Assistance Office of Economic Cooperation Bureau
	Evaluation to secure coherence/comprehensiveness	Promotion of receiving foreign students (2003)	Director for Policy Evaluation and Administrative Review, General Affairs Division of Minister's Secretariat, Consular Affairs Bureau, Cultural Affairs Department, Japan Foundation
	Evaluation to secure objectivity	Annual MOFA policy evaluation	Director for Policy Evaluation and Administrative Review, General Affairs Division of Minister's Secretariat
Administrative evaluation & inspection	International cultural exchange project		General Affairs Division of Minister's Secretariat, Cultural Affairs Department
	Administration of implementation system for diplomacy and overseas operations (MOFA Reform, 2003-04)		General Affairs Division of Minister's Secretariat, First Divisions of relevant Bureaus
	Promotion of e-government (2003)		General Affairs Division of Minister's Secretariat
Evaluation of IAAs	Japan International Cooperation Agency (annually)		General Affairs Division of Minister's Secretariat, Director for Policy Evaluation and Administrative Review, Technical Cooperation Division of Economic Cooperation Bureau
	Japan Foundation (annually)		General Affairs Division of Minister's Secretariat, Director for Policy Evaluation and Administrative Review, Cultural Affairs Department

* To assess policy impacts, policy evaluation, administrative evaluation, and inspection are coordinated (Article 18, GPEA).

** Called Cultural Affairs Department at the time.

3. Meaning of policy evaluation

(1) Four types of policy evaluation

In Japan “policy evaluation” can have four different meanings which causes confusion because the differences in meaning are not easy to distinguish or understand.

The first category is the “policy evaluation” which simply combines the

colloquial term “policy” with another colloquial term “evaluation.” Nippon Keidanren (Japan Business Federation) provides the most well-known example of a policy evaluation. In its policy evaluation, Keidanren presents the directions it deems desirable with regards to issues such as taxation, and social security. It also evaluates and rates the stated commitments, actions in Diet deliberations, and legislative performance of each political party. Keidanren member companies use it as a political party “report card” for deciding which political party to make contributions to and how much those contributions should be. In 2003 Keidanren launched this evaluation on a pilot basis, and in 2004, they rated the policies of political parties using five grades (A-E). Also, Genron NPO’s evaluations of manifestos, policies and administrations are evaluations of policy results in a broad, general sense, and can be categorized as “policy evaluation.”¹⁾

In addition, in more than a few cases policy evaluations are conducted without the knowledge of the parties involved. A typical example is a newspaper article based on a reporter’s own field investigation of a government policy change. In the spring of 2007, for example, Asahi Shimbun investigated the “New Program to Stabilize Farmers’ Income” launched by the Ministry of Agriculture, Forestry and Fisheries (MAFF) (Asahi Shimbun Kyoto Edition, December 2, 2007). The Asahi Shimbun’s investigation (evaluation) revealed that contrary to MAFF’s intentions, changes it made to the subsidy system caused a decrease in wheat farmers’ income in Hokkaido and Fukuoka prefectures.

The second category includes various ministries’ reviews, studies, and researches. Specific examples include the comprehensive ministry policy reviews conducted by the policy-related bureau of each ministry. Policy-related bureaus include: MOFA’s Foreign Policy Bureau; the Policy Bureau of the Ministry of Land, Infrastructure, Transport and Tourism (MLIT); and the Environmental Policy Bureau of the Ministry of the Environment (MOE). The Foreign Policy Evaluation Panel, for example, used to review foreign policies from a long-term perspective, separately from the organizations in charge of policy evaluation (the General Affairs Division and the Office of the Director for Policy Evaluation and Administrative Review) established within the Minister’s Secretariat based on GPEA. (The Foreign Policy Evaluation Panel was established within MOFA’s Foreign Policy Bureau and performed reviews from August 2002 to September 2003.) In addition, MLIT’s initiative in regards to “the review year for each law related to regulations” is noteworthy in this second category. Under MLIT’s initiative, each law is supposed

to be reviewed every 5 or 10 years, and this review is in itself a policy evaluation ²⁾. Of course, evaluations of development assistance “policy,” especially policy- and program-level evaluations also fall into this second category.

The third category includes examples of reviews from a “policy-related perspective” by a particular ministry. A specific example of such a review in this third category is a study by the Gender Equality Bureau of the Cabinet Office conducted on the policies of other ministries. This bureau, which does not make or implement actual policies, looked into the methods of “impact assessments” in a study written by the Impact Assessment Working Team of the Gender Equality Bureau of the Cabinet Office titled, *Impact Assessment Case Study Working Team’s Interim Report: A Trial of Assessment Methods for Planning and Implementing Measures from the Perspective of Gender Equality* (November 2003). A similar example of a study in this third category is the *Perspective for Considering the Comprehensive Inspection of Public Administration* by the General Planning Subcommittee of the Social Policy Council of the Cabinet Office (November 26, 2007) ³⁾.

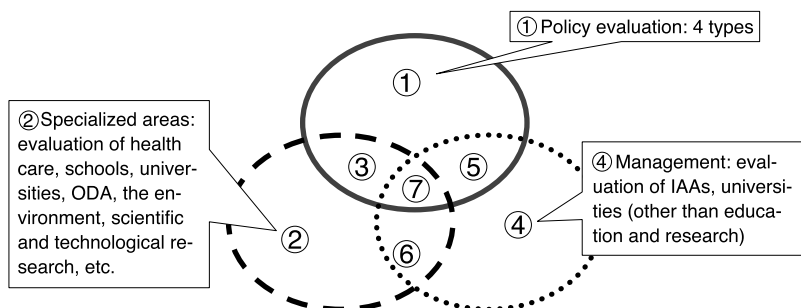
A fourth and last category includes policy evaluations conducted under the GPEA. Typically, the Accounting Division (budget control) and the Personnel Division (personnel and positions management) of the Minister’s Secretariat collaboratively conduct these administrative management-type policy evaluations. Despite being annual “evaluations” of “policies,” the definitions of “policy” and “evaluation” are quite different from the definitions used by academics and professionals in the fields of education, health care, and ODA. In addition to sometimes eliciting quizzical looks for their choice of definitions, the system is a bit confusing because of the coexistence of three different modes of evaluation with different purposes and methods: “comprehensive evaluation,” “performance evaluation,” and “project evaluation.”

(2) Clarifying the concept of “evaluation”

In general, “evaluation” includes the following five techniques. Two of these five techniques, “measurement” and “evaluation” comprise the primary tools of evaluation and are beginning to be recognized by the new term “M&E” ⁴⁾.

- Evaluation: To “study” the evaluation subject. If results are not achieved, “think” about the causes. The “comprehensive evaluation” under the GPEA, for example. In MLIT it is called a “review.”
- Analysis: To “divide and think” about costs and results. The “project evaluation” under the GPEA, for example.

Figure 2. Areas where evaluation and similar “evaluation” type activities take place



- **Measurement:** To “measure” the evaluation subject (basically the performance output). “Performance evaluation” under the GPEA, and performance measurement in the evaluation of IAAs, for example.
- **Benchmarking:** “Compare” the program’s outcome and performance with successful or previous cases.
- **Research:** Extensive, in-depth, time-consuming study.

On the policy frontline where suitable techniques are selected from these five, the way they appear and the way they are seen are somewhat complex. Figure 2 shows them in a simplified diagram.

The three basic areas where evaluation and similar “evaluation” type activities take place are represented by ①, ② and ④ in Figure 2. They are considered below in the order that they appeared in the history and evolution of evaluation ⁵⁾.

Evaluation in specialized areas underlying policies (Area ② in Figure 2), such as education, school, environmental, and health and welfare evaluations appeared first. Disciplines strongly related to the respective professions (education, forestry, environmental assessment, medical and health science) are deeply involved in these evaluations. They are sometimes referred to as researches or reviews.

Evaluation of policies themselves (Area ①: policy evaluation) appeared next. These policy evaluations are conducted by the bureaus, divisions, or officers of the central government’s ministries whose titles include the word “policy” in them. The review of management (Area ④: management review) appeared next. In the central ministries, the so-called “three divisions in the Minister’s Secretariat” (General Affairs Division, Accounting Division, and Personnel Division) conduct management reviews. Most of the subjects of

these reviews are organizations at the field level or outpost agencies.

Assessments in the specialized areas (Area ②) began as evaluation studies and transitioned to evaluation researches and professional evaluations. These subsequently developed into program evaluations or public works-related project evaluations where experts performing cost benefit analyses on projects' impacts (Area ③), which finally led to policy evaluations activities or more generalized "evaluations of policies" and policy reviews (Area ①). In the general society of Japan, this policy evaluation is conducted as four different types of "policy evaluations" as described above in Section 3 (1).

In addition, the area of management review is influenced by the concepts of New Public Management (NPM) and "re-inventing government." Under the influence of these concepts, evaluations are sometimes used as performance measurements in the institutional framework of IAAs and universities, or as information on the progress of marketization. Sometimes management review (Area ④) exclusively deals with evaluations of IAAs – i.e., assessment of progress toward achieving medium-term targets, operational performance evaluations during the medium-term target period, and annual evaluations – but it can also deal with project evaluations of IAAs and measurements of project performance. By the way, MOFA's objectives in regards to its Plan for Increasing the Efficiency of Public Administration (Cabinet Secretariat, Liaison Meeting of Related Ministries for Increasing the Efficiency of Public Administration, February 5, 2004), are not very different from those of IAAs.

As described above, the basic three pillars are clear. However, in the real world, where problems are complex, actual evaluation activities often involve the three basic areas overlapping and interlacing one another. Specifically, with regard to program evaluation or public works-related project evaluations (Area ③), there are examples where evaluation studies and researches, which used to be conducted in specialized areas, were utilized as a means to pursue the accountability of government programs and as "program evaluations." Program evaluations are often used because they evaluate social programs that provide services to people relating to education, employment, health, and welfare.

There are two examples of situations that deal with "measurements" rather than evaluations as represented by Area ⑤ where policy (Area ①) and management (Area ④) overlap. One example is the "performance evaluation" used as a mode of policy evaluation, which in reality consists of quantitative measurements designed to assess the outcomes and achievement sta-

tus of policy targets set by the executive branch or the Diet. The second example is the measurement of “operational performance” of “implementing agencies” charged with the means to implement policies, such as IAAs and private contractors. The subjects of these measurements are usually outputs of activities rather than outcomes. The essence of both the performance “evaluation” and operational performance “evaluation” is really performance “measurement,” and therefore differentiating between performance evaluation and operational performance evaluation is difficult. In addition, they are different in character from evaluations, but this fact is difficult for nonexperts to understand.

In Area ⑥, officials use this method of “performance measurement” and try to quantitatively assess activities in specialized areas. They use a method related to “management by objectives” and a method of establishing certain criteria and counting the number of organizations that meet the criteria. For example, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) has established a competitive financing system modeled after the Center of Excellence and Good Practice (COE & GP) to provide policy incentives to universities and graduate schools. Universities that meet more criteria are placed higher in university rankings and receive favorable treatment ⁶⁾. Private universities are sometimes treated differently and rated on four criteria: “trend of tuition revenue,” “trend of non-tuition revenue,” “stock” and “governance and management.” This method is also designed to foster sound institutions of higher education.

The problem lies in the situation represented by Area ⑦, where all those areas overlap. Theoretically, it deals with “comprehensive evaluations” used for the judgment of top managers of organizations or elected officials (although they are different from the mode of “comprehensive evaluation” used under the Evaluation Law ⁷⁾). It is an evaluation conducted for the purpose of obtaining information that allows a review of a “high policy” which involves a high-level “policy judgment” from a long-term perspective that spans 10 to 20 years. Examples include the comprehensive review and recommendation on foreign policy by MOFA’s Foreign Policy Evaluation Panel (see above) ⁸⁾, and studies such as “The Twentieth-Year Review of Japanese Structural Adjustment Loans ⁹⁾.” Since it examines and re-examines a high policy from a broad, long-term perspective and considers the validity of institutional building in light of policy objectives, we can also say that it is a strategic management tool ¹⁰⁾.

In reality, however, officials on the ground are not aware of the complexi-

ties as illustrated by the Areas ① through ⑦ when they construct evaluation systems, select data gathering tools or decide on the method for analyzing and comparing evaluation systems. Even if they could distinguish between the different types of evaluations and want to use them appropriately for different purposes, the process would take too much time and trouble. Therefore, there is demand at the practitioner's level for a set of selection criteria that could be a simple checklist to use when choosing an evaluation method. In fact, two such methods are being tried on the ground. The first is Article 10 of the GPEA that stipulates,

“When the head of an administrative agency conducts a policy evaluation, he/she shall prepare an evaluation report which describes the following”:

1. Evaluated policy or policies
2. The division or organization that performed the policy evaluation and when it was performed
3. Perspective of the policy evaluation
4. Methods and results of policy impact assessment
5. Matters related to utilizing people of experience and academic standing
6. Matters related to data and other information used in the process of policy evaluation
7. Results of the policy evaluation

The practitioner must think about these seven items and use them as a checklist when he/she performs an evaluation.

The second method is to establish an organization consisting of outsiders that checks the quality of evaluations. Specific examples of this method include; establishing an expert committee consisting of external experts; and having the committee evaluate the evaluations themselves (meta-evaluation). External experts, who should be appointed in a balanced manner, may include policy experts (who are familiar with the policy-making process of the government), experts on administrative management and evaluation (including certified accountants, management consultants and corporate executives), and professionals in the specialized area behind the policy in question (educationists, doctors, etc.).

4. ODA evaluation and policy evaluation

Now, after having spent more than ample space on a general discussion of evaluations, the remaining portion of this chapter is devoted to the main topic

of explaining the policy evaluation and the ODA evaluation.

In principle, there is no particular reason to distinguish ODA evaluations from evaluations in general, and therefore, there are no differences between them. In 1979, three famous scholars of evaluation studies, Freeman, Rossi, and Wright were invited by the OECD to give speeches and provide training on evaluation right after they published, *Evaluation: A Systematic Approach* which was the first textbook of its kind ¹¹⁾. The OECD, subsequently published manuals and guidelines on various types of evaluations including policy evaluation, program evaluation, and performance measurement. The OECD does not distinguish ODA evaluations from policy evaluations. In addition, since aid recipient countries share common evaluation criteria with aid donor countries such as the realization of outcomes and efficiency, it is not particularly meaningful to distinguish between the evaluation of its own policies and that of foreign aid. Nevertheless, we tend to have a false impression that there are in fact differences between policy evaluation and ODA evaluation and there are some reasons for this misconception.

One reason for the misconception can be found in the long history that created a unique discipline of ODA evaluation. The history has its roots in 1975 when the former Overseas Economic Cooperation Fund (latter-day Japan Bank for International Cooperation) began to address the evaluation issue and in 1981 when MOFA and JICA officially began to perform evaluations. The discipline of ODA evaluations was built by the practitioners of foreign policy and ODA, as well as scholars of international relations, international economy, international finance, and regional studies. Compared to the disciplines of public administration and public policy, the history of ODA evaluation took shape in a different world with a different way of thinking underlying discussions. (In fact except in the field of “development administration” in the U.S., public administration and public policy rarely discussed ODA evaluation). Naturally, a peculiar jargon was created and has grown. When ODA became the subject for the “Administrative Inspection on Economic Cooperation Part 1: Grant Aid and Technical Cooperation” in 1987 and for “Part 2: Loan Aid” in 1988; and the Foreign Affairs Audit Division was created in the Board of Audit in 1998 ¹²⁾, a small window opened, for scholars of public administration (who were outsiders at the time), allowing a peek into this world of international development and ODA.

Another reason for the misconception is the perception gap between foreign policy, considered “high policy,” and administrative control. In the mid-1990s when policy evaluation attracted attention and was introduced on a

pilot basis, ODA evaluation had already been “institutionalized” to a certain degree; i.e., theories on the practice of foreign policy had been established and evaluation theories had been applied to the practice of development. On the other hand, in a separate world from ODA evaluation and foreign policy, research on the theories and practice of policy evaluation progressed in a fumbling manner. Subsequently, however, its direction was changed by the request of the Council on Economic and Fiscal Policy and moved toward the “administrative management-type policy evaluation,” the result of which are reflected in budgetary assessments and positions management. For this reason, foreign policy experts in charge of ODA evaluation felt uncomfortable toward policy evaluation. A typical example of this discomfort was found with regard to the ex-ante evaluation of projects (Article 5, Section 4 of the GPEA). There was a general feeling within MOFA’s Economic Cooperation Bureau that, while ex-ante evaluations of public works projects are naturally required for the ministries in charge of domestic public works, ex-ante evaluations are not suitable for ODA because ODA is related to foreign policy. Subsequently, however, the ex-ante evaluation of projects was implemented after a one-year research period set by a joint MOFA-MIC ministerial decree.

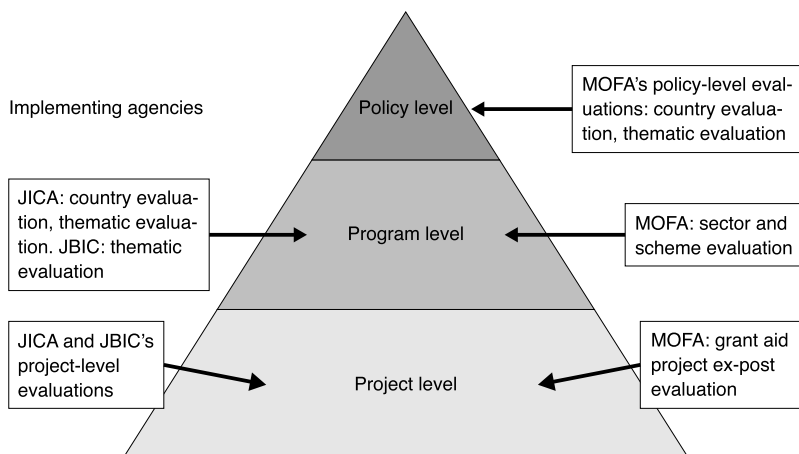
To begin with, as we can see in Figure 2, the professional evaluation of development assistance and ODA, which belongs to Area ②, and the administrative management-type policy evaluation and evaluation of IAAs in Areas ④ and ⑤ reside in different worlds (JICA was reorganized as an incorporated administrative agency in 2003). However, neither MOFA nor JICA could ignore the policy evaluation and the evaluation of IAAs, which were institutionalized by legislation favoring management as influenced by NPM, and both entities are now conducting evaluations within the framework of this system. Then, where is the ODA evaluation positioned within the policy evaluation system, and what are the characteristics of ODA evaluation? Figure 2 is, again, useful to explain the different functions of ODA evaluation and policy evaluation.

First of all, policy evaluations conducted by MOFA in Area ① and evaluations by JICA/JBIC in Area ④ are divided into different categories, because the roles of MOFA and IAAs are different (This is the difference between the policy evaluation conducted by MOFA proper and “corporation evaluations” of IAAs). However, there are two types of ODA-related policy evaluations conducted by MOFA. One is the “Ministry of Foreign Affairs Policy Evaluation,” conducted annually based on the stipulations of the GPEA and published as a booklet. The other is the “policy-level evaluation” within the

system of ODA evaluation. The former is drafted by different bureaus and divisions, and edited by the Minister's Secretariat (General Affairs Division and Director for Policy Evaluation and Administrative Review), while the latter is handled by the former Economic Cooperation Bureau which is now called the International Cooperation Bureau. With regard to evaluation activities under the GPEA, MOFA is required under Article 9 of the GPEA to conduct ex-ante evaluations of ODA projects (grant aid cooperation projects that cost 1 billion yen or more, and loan aid projects that cost 15 billion yen or more). MIC is supposed to implement the evaluation, "to secure coherence/comprehensiveness," based on Article 12, Section 1 of the GPEA, and as described above, economic cooperation (ODA) was its subject from 2001 until 2003¹³⁾.

In the context of ODA evaluation, evaluation activities are categorized into Areas ①, ② and ③. Policy-level evaluations represented by Area ① include country evaluations and thematic evaluations; program-level evaluations represented by Area ③ include sector evaluations and scheme evaluations; and project-level evaluations represented by Area ② include the ex-post evaluations of grant aid cooperation, which was launched in 2005. In FY2000, a study of policy- and program-level evaluations was conducted, and a system of division of labor was introduced in FY2001. Under this system, JICA conducts program-level evaluations (Area ③ in Figure 2: Evaluations of programs that cut across aid modalities and sectors and incorporate projects strategically to solve problems) and individual project evaluations (Area ②). JBIC conducts program evaluations that analyze the contribution of its financial assistance to the resolution of certain issues (Area ③) and project-level evaluations that analyze the impact of its financial assistance on specific infrastructure-building projects (Area ②). In this division of labor system, policy-level evaluations are handled by MOFA; program-level evaluations by MOFA, JICA, and JBIC; and project-level evaluations by JICA and JBIC (See Figure 3).

This hierarchical division of labor system for "policy/program/project" is called a policy structure, and is commonly seen not only in the field of ODA but generally found in the world of evaluation. With regard to evaluations conducted by national and local governments, it is structured as a demarcation system of "policy/program/administrative project (project)", with a firm understanding of each category as a unit of evaluation. However, as we discussed at the beginning of this chapter, these concepts themselves are ambiguous and create a confusing situation where some administrative pro-

Figure 3. Division of labor system for ODA evaluations

ject evaluations are called administrative evaluations. This situation makes coherent discussions difficult.

One of the underlying causes is the lack of progress in decentralization. Prefectures and municipalities (cities, towns, and villages) are regarded as implementing agencies of the central ministries, but a true transfer of power and authority has not taken place. For this reason, local governments (especially municipalities) that cannot transform themselves into policy agencies do not need to produce policy data to be used for policy selection. Currently, as there remains a framework of centralized division of labor where policies and programs are handled by the central ministries and projects are handled by local governments, local governments only need administrative project evaluation because they de facto project implementing agencies. Moreover, the tight fiscal conditions of local governments do not allow them the luxury of implementing policy evaluations to assess the impacts of policies before constructing new strategies. Instead, the situation forces local governments to adopt the evaluation of public administration management that emphasizes reductions and savings above all other criteria. Paradoxically, the centralized system has created a “policy/program/project” structure in the relationship between the central and local governments. In this structure, the central government handles policy evaluations and local governments handle administrative project evaluations, which remind us of the relationship between the central ministries and IAAs. Within this structure, a central ministry sometimes asks a local government to evaluate projects under its policy jurisdic-

tion, and to evaluate subsidy projects as a condition for receiving subsidies. The latter is similar to the system of “conditionality” in ODA.

5. Problems of ODA-related policy evaluation in Japan

In Japan, policy evaluations are based on the image of “high policy” evaluation, but, in reality, policy evaluations have turned into administrative management-type evaluations. With an awareness of this situation, let us think again about the relationship between policy evaluation and ODA evaluation.

In the beginning, policy evaluation and program evaluation differed from performance measurements of NPM in terms of their origins and activities. However, because of the ambiguous application of the concept of “evaluation,” and because they were all introduced from overseas approximately at the same time, policy evaluations and performance measurements were adopted without a clear understanding of their differences. A prominent manifestation of this is the so-called “administrative evaluation” of local governments. ODA evaluation, on the other hand, is also influenced by NPM. For example, since the concept of “Result-Based Management” and methodology of “M & E (measurement and evaluation)” are both incorporated, ODA evaluation is, on the surface, similar to administrative evaluation and policy evaluation. Some may say that ODA evaluation is not different from policy evaluation in Japan. Nevertheless, there is no doubt that, in reality, their activities are quite different.

Human resources and how human resources are deployed stands out as a real difference. Policy evaluation, in particular, lags behind ODA evaluation in terms of the quality and quantity of evaluators. In each ministry, a small number of evaluators have to devise and manage the system, methods, and schedules of evaluation, which makes maintaining the quality of evaluations difficult. At every regular personnel reshuffle, a complete novice assumes the post of an evaluator. This causes a gradual dilution of the “evaluation mind” which leads to evaluation reports without analyses, and at the same time, difficult questions such as: “I would like to know an evaluation method to assess the impacts of financial contributions to an international organization, the contribution of shares of which are pre-determined by an international agreement.” Perhaps we should improve the evaluation mind by; training, researches and studies; and implementing “capacity building” activities as is done for ODA evaluation. However, from the perspective of evaluators in each ministry, few training courses are provided by instructors who are both

familiar with the ministry's policy-making processes and who have no organizational bias. In addition, in ministries where evaluators are assigned to high-level posts, they cannot focus exclusively on policy evaluations (or evaluation of IAAs), and are often given other assignments, which sometimes relegate evaluations to the status of a side job. Sometimes, the posts of evaluators are used as mere points of passage or "temporary waiting places" for career bureaucrats.

The second difference between policy evaluation and ODA evaluation lies in the response to the demands for accountability. Institutionally, there is a flow of accountability with regard to policy evaluation: Policy Division in each ministry → organization in charge of policy evaluation in each ministry → external expert committee on policy evaluation in each ministry → MIC's Administrative Evaluation Bureau → MIC's Committee on Policy Evaluation and Evaluation of IAAs → the Cabinet → the Diet. In this mechanism, accountability for policy evaluation itself is ensured by properly following the procedures and by the institutional framework established in the governance system through which the Cabinet and the Diet demand accountability.

On the other hand, with regard to ODA evaluation, it is difficult for an outsider to directly confirm an institutionalized political accountability mechanism other than the processes and procedures described above. Even if it exists, it is fairly complex and not as simple as in domestic policy evaluation. As a result, procedures for, and the contents of accountability are also different. Accountability in this context is accountability from a technical perspective. Therefore, the requirement to "use objective third parties" is translated into "external experts," and the validity of evaluation depends on whether it is rational (explainable) in the eyes of experts. MOFA; JICA; JBIC; FASID; various think tanks, universities and graduate schools involved in international development and international relations; the Japan Evaluation Society; and Japan Society for International Development often hold study meetings and training sessions because the professionals in these organizations hope to develop and maintain their abilities in order to fulfill their obligations as experts in regards to accountability and because they hope to qualitatively improve ODA evaluations. Although these efforts appear to be accountability, in fact, they do not represent true accountability because for true accountability, responsibility is monitored (or governed) by outside authorities and under the threat of potential sanctions. Rather, they reflect a responsibility backed by the abilities and self-pride of professionals to work proactively with a sense of duty. However, even with regard to domestic policy evalua-

tion, the primary evaluation begins with a self-evaluation of an official in charge of a policy, and the administrative monitoring function of the Diet, which should ultimately check the quality of that evaluation, is not coordinated with this part of the evaluation. MIC's reviews of ministries' policy evaluations seem like technical advice, and those who fill in evaluation sheets or write evaluation reports are not held accountable by the general public or the Diet. In this situation, the pursuit of accountability itself is weak, and politicians are not setting directions. Even the domestic policy evaluation is not really an effective tool for ensuring accountability.

The third difference between policy evaluation and ODA evaluation lies in their nature. In ODA evaluations, too, ex-ante projections and ex-post verifications are conducted within a certain framework. However, a large part of the evaluation itself is not necessarily incorporated into the budgetary assessments by MOFA and the Ministry of Finance (MOF) or into the positions assessments by MIC's Administrative Management Bureau. Also, evaluations described in MOFA's *Annual Evaluation Report on Japan's Economic Cooperation*, JICA's *Annual Evaluation Report* and JBIC's *Yen Loans Annual Evaluation Report* are mostly professional evaluations (Area ② in Figure 2) that primarily emphasize professional accountability to the outside world. On the other hand, those involved in the domestic policy evaluation became acutely aware of budgetary management and positions management in 2005, and thought about ex-ante assessments and ex-post verifications (especially the document submitted by the then-Finance Minister Tanigaki to the Council on Economic and Fiscal Policy on March 10, 2005). Here, they were trying to find a way to bring the "line items" of budgets and financial statements closer to "programs," and were considering a transition from "program performance measurement" to "program budgeting." This is why in this chapter we coined the term "administrative management-type policy evaluation." They have tried, through structural reform, to realize the intention for the General Affairs Division, Accounting Division, and Personnel Division of the Minister's Secretariat to be in charge of budgetary, positions (personnel) management, and utilize evaluation data. In other words, in terms of Figure 2, a shift from Area ① to Areas ⑤ and ④ is occurring within policy evaluation. In extreme cases, we are beginning to see situations where it is impossible to distinguish a ministerial policy evaluation from a project or performance evaluation of an incorporated administrative agency. Thus, policy evaluation began to acquire an appearance of management evaluation, while ODA evaluation began to acquire an appearance of professional

evaluation.

The fourth difference is related to the complexity of evaluation systems. Since policy evaluation required a consensus (or lack of objection) among the officials in charge in each ministry, it excluded details that were difficult for the ministries to agree upon. For this reason, policy evaluation remained relatively simple. The three basic evaluation modes proposed by MIC and agreed upon by other ministries were the comprehensive evaluation mode, the performance evaluation mode, and the operational performance evaluation mode. How to choose an evaluation mode was supposedly left to each ministry. However, since MIC's Administrative Evaluation Bureau provided "technical guidance" on the results of the selected evaluation modes, ministries' evaluations converged in a certain direction. In addition, evaluations were basically scheduled after completion (ex-post). In the GPEA, ex-ante evaluations are required only in exceptional areas explicitly limited to; research and development; public works; and ODA.

In contrast, ODA evaluation is complex and year after year, it has become even more complex, sophisticated, and specialized because of the conscience of experts, the public's critical watch over ODA spending, and the expansion and diversification of ODA activities. At the same time, although evaluation theory itself has become more complex, the theory has not been able to freely influence the practice of evaluation. Generally, they analyze new requirements as they come, and select the items in Table 2 one by one to design an evaluation that meets these requirements.

For example, the Council on the Proper Implementation of Grant Aid Cooperation was established in response to a demand for better implementation of grant aid cooperation. This evaluation institution was created as a result of selecting appropriate items based on requirements, from Table 2 above, the ex-post evaluation (timing), meta-evaluation (evaluation body), external experts (internal/external), projects (subject), and grant aid cooperation (scheme). Therefore, it is likely that new evaluation systems will emerge in the future with different combinations of items in Table 2. And, consequently, we will end up with a large number of evaluation systems that are so complex and highly specialized that non-experts will have a hard time understanding them. In this scenario, it will be quite difficult to ensure accountability in terms of the "ability to explain and convince others." There is a dynamic in ODA evaluation that works at a different level than that of the relatively simple policy evaluation. However, when the complexity increases beyond a certain level, the situation will become uncontrollable, and there will be a

Table 2. Various type of ODA Evaluation

Item	Content
Timing	Ex-ante, mid-term, at completion, ex-post (more than 3-5 years after completion)
Evaluating body	Primary evaluation, secondary evaluation, meta-evaluation, joint evaluation
Internal/external	Internal, external (commission, experts, NPO)
Subject	Policy, program, project, organization, discipline
Scheme	Grant aid, loan aid, technical cooperation, grass-roots grants, cultural grants
Purpose	To strengthen the check function; greater accountability to the public

need for reorganizing and taking stock of various evaluation systems. “Proposals for reforming the evaluation system” and “task forces for improving the evaluation system” that sometimes pop up in MOFA are attempts to reorganize the modes of evaluation that have become too complex to understand in terms of the evaluating body, subject, evaluation method, feedback and publication method. This may be a characteristic that is not seen in policy evaluation.

Finally, I would like to point out that there is one, exceptional area where ODA evaluation and policy evaluation overlap. It is the description of each ministry’s policy evaluations in MOFA’s ODA evaluation (*Annual Evaluation Report on Japan’s Economic Cooperation*). This section includes both evaluations based on the GPEA and evaluations based on the judgment of each ministry. However, this area has nothing to do with the discussion of this chapter: This area exists simply because MOFA is the ministry that coordinates relevant ministries on matters related to ODA.

6. Concluding remarks

The starting point of our discussion in this chapter was the fact that both ODA evaluation and policy evaluation contain an aspect of policy evaluation. In discussing the differences between ODA evaluation and policy evaluation, our discussion unwittingly turned into a comparison of the practice and theory of ODA evaluation with those of domestic, administrative evaluation.

In general, domestic policy practices that become the subjects of political science and public administration studies are; planning of policies and selection of policy instruments (subsidies, financing, regulations, deregulations, taxation, public relations, education, etc.) at the central ministries; building of

and changes in the implementing systems within the IAAs and local governments responsible for implementing these policy instruments (privatization, marketization, merger, decentralization, delegation of authority); and evaluations that reflect on these activities. Of course, we also think about how “politics” influences them, and what kinds of restrictions economic conditions impose on the planning and selection of policies. (Of course, the policy evaluation and evaluation of IAAs conducted by MOFA, one of the central ministries, should be included in this discussion, but this has not been the case due to the excessive sense of “division of labor” among domestic researchers.)

On the other hand, in the practice of ODA, just as in the practice of domestic policies, projects are identified, formulated, implemented and evaluated with help from experts in the related fields such as education, health care, and agriculture. However, while these two practices are similar, they have a subtle misalignment: The fact that the ODA evaluation considers the meaning of ODA in diplomacy is a clear manifestation of that misalignment. For example, both ODA evaluation and policy evaluation consider; “policy” areas such as education, health care and agriculture; “professions” in the respective fields (teachers, educationists, doctors, health care researchers, agricultural trainers, etc.); and related “disciplines” (education studies, medical science and health care studies, agricultural studies, etc.), and each of them has a different perspective on evaluation. For example, there is a subtle difference between an evaluation by an agricultural organization of a donor country that concludes, “the policy was successful and achieved targets efficiently,” and a local agricultural expert deciding that agriculture did well in the aid recipient country. In addition, sometimes management pressure is applied. Management pressure with regard to ODA was applied during JICA’s reorganization as an incorporated administrative agency (October 2003), and with regard to policy, it manifested itself in the shift of policy evaluation towards “administrative management-type policy evaluation.” A judgment of success at the management level as a result of an evaluation, a perception that a policy was a success, and a judgment that agriculture was sustainable do not belong in the same dimension. What we discussed so far can be seen both in ODA evaluation and in domestic policy evaluation. However, a discussion of whether a policy contributed to Japan’s foreign policy is also a story at a different level. To put it plainly, when policy-level issues are considered in an ODA evaluation, more issues must be considered than in a policy evaluation of domestic public administration, and the level of difficulty for a

policy-level ODA evaluation is significantly greater than that of a policy-level evaluation of domestic public administration.

However, there are different kinds of difficulties associated with domestic policy evaluation. Domestically, the influence of “politics” and vested interests reaches even very detailed areas that non-experts are not even aware of. For this reason, it is not easy to bring changes to subsidies, earmarked revenue sources, the taxation system, and regulations. Unlike in the field of ODA, the freedom in the selection of policy instruments is quite restricted on the domestic front. Very strong political leadership is necessary to overcome the resistance, and, as we experienced during the privatization of the postal service, we need to prepare ourselves for ideological changes in politics and political parties including the governing party.

In addition, there are values that have been established through the long history and researches in domestic policy areas. For example, pairs of terms in use since the post-war democratization era that appear similar in usage, have subtle differences. A few examples include: local government and local public institution: self-governance and local administration: citizen and resident (with regard to participation and activism): delegation of authority and devolution. If you make a poor choice of words, you may be laughed at or ignored by the opposing side. (This atmosphere may be reflected in the pair of similar terms “administrative evaluation” and “policy evaluation” as well.) Furthermore, academic disciplines have been established based on such nervous discussions, and corresponding professions have been created. Therefore, in fact, a “world” similar to that of ODA has been created domestically.

Because people in different disciplines and professions used the same word “evaluation” in separate worlds, lay people who are not academic researchers or professionals have the wrong impression that “evaluation” in these two different contexts means the same thing. However, in each context, the word “evaluation” has a slightly different meaning, and in many cases, the definition of evaluation in one context is not compatible in another context. The word “evaluation” can mean quite different things in different contexts.

However, since the trend of the world and the times is moving towards administrative management-type policy evaluation, both ODA evaluation and domestic policy evaluation will, despite some resistance, eventually converge in this direction. This is the conclusion of this chapter. However, we need to note that, unfortunately, the application of administrative management-type

policy evaluation is limited. In that sense, MOFA's *Policy Evaluation Report*, *Annual Evaluation Report on Japan's Economic Cooperation*, *White Paper on Official Development Assistance*, JICA's *Annual Evaluation Report*, and JBIC's *Yen Loans Annual Evaluation Report* are without a doubt very important sources of intelligence that is necessary when one wants to look at ODA policies in a comprehensive manner. What we need to do is to develop competent people who can recognize the diversity of evaluation and respond to the needs of the times with a cool head. Interdisciplinary researches and education are essential to that end, but it is a very difficult task.

(*This article is translated from "ODA hyoka to seisaku hyoka: Nihon no genjo bunseki," *Kaihatsu enjo no hyoka to sono kadai*, Kaihatsu enjo doko series, 2008, FASID.)

Notes

- 1) For Genron NPO, established by Yasushi Kudo in November 2, 2001, see its website (<http://www.genron-npo.net/>).
- 2) See the website of the Ministry of Land, Infrastructure, Transport and Tourism (<http://www.mlit.go.jp/hourei/itirann.pdf>). The way this initiative works is very similar to that of the so-called sunset laws adopted by various states in the U.S. in the late 1970s.
- 3) See the website of the General Planning Subcommittee of the 21st Social Policy Council of the Cabinet Office (<http://www5.cao.go.jp/seikatsu/shingikai/kikaku/21th/index.html>).
- 4) Cf. Keith Mackay, *How to Build M&E Systems to Support Better Government*, World Bank, 2007; James C. McDavid and Laura L. Hawthorn, *Program Evaluation & Performance Measurement: An introduction to Practice*, Sage, 2006.
- 5) Evaluations began as "professional" evaluations of various social programs such as education, welfare and health care (the so-called "evaluation research"), and this method was used for the purpose of ensuring the government's accountability in an attempt to evaluate "policy" programs ("program evaluation"), with an additional application as a "management" tool for among others, monitor managers who engage in policy implementation, and measure outputs and outcome indicators. In the late 1980s when a discussion about using evaluation for management began, NPM appeared and influenced policy evaluation. For these developments, see Kiyoshi Yamaya, *Theory of Policy Evaluation and Its*

Development (Koyo Shobo, 1997), Chapter 3.

- 6) Examples of the COE & GP-type competitive financing include the Program for Promoting High-Quality University Education, Program for Promotion of Education Responding to Adult Reeducation Needs, Student Support Program Responding to New Social Needs, University Education Internationalization Promotion Program, Program for Promotion of Education for Developing Highly Specialized Professionals in Professional Graduate Schools, Global COE Program, Graduate School Educational Reform Support Program, Strategic University Partnership Support Project, Project for Promoting the Development of Medical Professionals through Partnership with University Hospitals, Cancer Professional Training Program, Project for Practical Human Resource Development through Industry-Academia Collaboration, Leading IT Specialist Training Promotion Program.
- 7) Peter H. Rossi, Howard E. Freeman, and Sonia R. Wright, *Evaluation: A Systematic Approach*, 1979, p.16
- 8) On February 12, 2002, Foreign Minister Junko Kawaguchi announced the “Ten Reform Principles to Ensure an Open Ministry of Foreign Affairs,” proposing to make policy-making processes more transparent and to establish the Foreign Policy Evaluation Panel as a means to reflect the opinions of various sectors in the policies of MOFA. Also, the final report of the MOFA Reform Advisory Board, submitted on July 22, 2002, recommended to “strengthen the ability to envision policies” and to actively utilize “policy evaluations,” and more specifically, to establish an organization to conduct policy evaluations or “an external policy evaluation panel” within the Foreign Policy Bureau and “utilize the panel for policy-making with regard to strategic (medium- to long-term) foreign policy issues” (http://www.mofa.go.jp/mofaj/annai/honsho/kai_genjo/pdfs/hyoka_panel.pdf). We can see from this that the Foreign Policy Evaluation Panel was designed for comprehensive policy reviews and established by the initiative of the Minister as a high-level organization within MOFA. Also, we can understand that the policy evaluation is of a different type from the “administrative management-type” policy evaluation conducted in collaboration with the so-called “three divisions in the Minister’s Secretariat” (General Affairs Division, Accounting Division and Personnel Division).
- 9) Yayoi Tanaka “The Japanese Government’s Policy and Judgment as Seen in the Twentieth-Year Review of Japanese Structural Adjustment Loans,”

The Japanese Journal of Evaluation Studies, Japan Evaluation Society, Vol.6, No.1, March 2006.

- 10) John Mayne, "Evaluation for Accountability: Myth or Reality? " Marie-Louise Bemelmans-Vides, Jeremy Lonsdale and Burt Perrin, ed., *Making Accountability Work*, Transaction Publishers, 2007, Chapter 4.
- 11) Kiyoshi Yamaya, *Theory of Policy Evaluation and Its Development - Accountability of the Government*, Koyo Shobo, 1979, p.129.
- 12) Kiyoshi Yamaya, "Theory and Practice of Evaluation in Official Development Assistance - Japanese ODA Projects," Institute of Administrative Management, *ODA Evaluation System - Theory and International Comparison*, March 1993, Chapter 2.
- 13) Ministry of Internal Affairs and Communications, *Policy Evaluation Report on Economic Cooperation (Official Development Assistance)*, April 2004.

Evaluation Capacity Development: A Practical Approach to Assistance

Takako Haraguchi, Keishi Miyazaki

1. Introduction

This chapter serves as a record of assistance in evaluation capacity development (ECD) through activities related to evaluation of Japan's Official Development Assistance (ODA) loan projects. Currently, ECD assistance for Japan's ODA loan projects is two-fold: (i) annual ODA loan project evaluation seminars and (ii) joint evaluation programs, both of which are hands-on ECD activities implemented by the Japan International Cooperation Agency (JICA) or the former Japan Bank for International Cooperation (JBIC) ¹. By outlining these activities and making observation based on our involvement in some of the actual activities, we present our thoughts in regards to developing countries' demands for ECD for ODA loan projects and how to respond to these demands.

The central issue we raise in this paper is: *how can developing countries come to feel the necessity to evaluate their development projects for their own sake?* A central issue of ECD is how to create the demand for establishing country evaluation systems ². The Paris Declaration promotes establishing country evaluation systems as a component of country-led development processes to enhance aid effectiveness. As such, the evaluation of ODA loan projects is a good entry point to raise the awareness and interest of developing countries to establish country evaluation systems.

Through our experiences as evaluation consultants and trainers for

1 The term "JBIC" is used in this report regarding events and activities before the merger of the former JBIC ODA loan departments and JICA in October 2008.

2 Keith Mackay, *How to Build M&E Systems to Support Better Government*, World Bank, 2007, pp.55-.

administrators of developing countries, we have learned that evaluation is a relatively new concept for many developing countries. Generally, demand for evaluation in developing countries has not yet been strong enough to encourage their governments to allocate sufficient resources for the establishment of the country evaluation systems. Developing countries often take a position that they just accept or follow initiatives by donors. However, as loan projects are generally larger (in terms of both costs and benefits) than technical assistance or other grant aid projects, and as developing countries must repay the loans using their own resources, improving those projects could be of higher interest of developing countries and thus evaluation of them.

Based on our experience and responses of participants that we will present in this chapter, we acknowledge that JBIC's current ECD assistance activities are quite effective. That is, *learning evaluation methods of international standard is a good starting point to generate interest in evaluation, and successive joint evaluations can further cultivate their demands for establishing country evaluation systems.*

In the following two sections, we will summarize and analyze our experience in ECD assistance. Section 2 describes the current ECD assistance model for Japan's ODA loan projects with some illustration of joint evaluation programs in Indonesia and Vietnam. Section 3 delves into two case studies comparing two joint evaluation programs in Vietnam, one supported by the Ministry of Foreign Affairs of Japan in 2005 and the other by JBIC in 2007. The two programs differed in terms of the demand for ECD on the part of the Vietnamese counterpart which in turn effected the degree of their participation in the program activities³.

Section 4 summarizes the key factors for enhancing and dealing with ECD demands. The most essential factors presented include (i) incorporation of components for institutional enhancement; (ii) involvement of project executing agencies in ECD; (iii) careful selection of evaluation methods and procedures to be transferred/ applied, including making adjustment in the Japanese ODA evaluation practices to align its evaluation methods to the evaluation methods of its partner countries; and (iv) good coordination among parties involved in evaluation.

3 Haraguchi has been involved in ODA Loan Project Evaluation Seminars since 2002 (Section 2) as well as some JBIC joint evaluation programs including the ones in Vietnam 2007 and 2008 (Section 3). Miyazaki was involved in all of the three joint evaluation programs in Vietnam 2005, 2007, and 2008 (Section 3). He is the team leader of the 2007 and 2008 joint evaluations.

2. A Model of ECD Assistance Through Evaluation of ODA Loan Projects

In this section, we outline the ECD assistance model JBIC has applied to ODA loan project evaluation in recent years. The model consists of the two stages: (i) ODA loan project evaluation seminars in Japan, followed by (ii) joint evaluation programs in the recipient countries. In the first stage, every year JBIC and JICA invite 15-20 partner countries to send representatives from their ODA planning/ coordinating agencies or executing agencies to a two-week evaluation seminar in Japan. Following the seminar, if a participant is interested in strengthening evaluation capacity in his/her organization or country, and if the concerned organization and JBIC can reach an agreement, JBIC starts the second stage: a bilateral collaboration program on ECD consisting of joint evaluation and other ECD-related activities⁴.

Table 1. JBIC's ECD Assistance Model for ODA Loan Projects

	Year 1	Year 2	Year 3	Year 4	Year 5
Phase	Needs identification	Technical transfer			Follow-up
Type of assistance	ODA loan project evaluation seminar	<ul style="list-style-type: none"> • Joint evaluation program (joint ex-post evaluation, institutional enhancement on evaluation) • Conclusion of collaboration agreement (Memorandum of Understanding) 			Follow-up of evaluation conducted by partner country themselves

Source: Adopted from Y. Wada, JBIC Challenges to Support Evaluation Capacity Development in Partner Countries, 7th Annual Conference of the Japan Evaluation Society, 2006, pp.227-230.

2-1 ODA Loan Project Evaluation Seminar in Japan

(1) Overview

Every year since 2001, JBIC and JICA have conducted a joint ODA Loan Project Evaluation Seminar (“the Evaluation Seminar”) in Japan. For each batch, 15-20 participants from developing countries are selected from government officers in charge of ODA planning/ coordination or project/ program planning, implementation, monitoring and/or evaluation (in principle one person from each country).

4 As illustrated in Section 3, the Ministry of Foreign Affairs of Japan also hosts annual “ODA Evaluation Workshop” inviting representatives from a number of countries. Vietnam’s joint evaluation program sprung from this in 2005.

Table 2. Outlines of ODA Loan Project Evaluation Seminars 2002 and 2007

	2002	2007
Seminar Objective	Transfer of knowledge of Japan's ODA project evaluation	Overall goal: evaluation capacity development of participating organizations/ countries through dissemination of knowledge acquired at the Seminar Immediate objective: evaluation capacity development of participants.
Participants	18 participants (Bangladesh, Bulgaria, China, India, Indonesia, Jamaica, Jordan, Kyrgyz, Laos, Mongolia, Morocco, Niger, Philippines, Thailand, Tunisia, Turkey, Vietnam)	17 participants (Thailand, Malaysia, Indonesia, Philippines, Vietnam, Cambodia, Sri Lanka, Albania, Pakistan, Turkey, India, Maldives, Egypt, Peru, Tajikistan, Tunisia)
Instructors	JICA/ JBIC Consultant (evaluation specialist)	JICA/ JBIC University International Organization Consultant (evaluation specialist)
Main Program	<u>Total 19 days</u> <u>Evaluation training</u> (Lectures and group exercises) - Project management of ODA loans and technical assistance - Project cycle management (PCM) planning method - PCM monitoring and evaluation method - Economic and financial analysis - Case study exercise on technical cooperation project evaluation - Case study exercise on ODA loan project evaluation	<u>Total 13 days</u> <u>I. Evaluation training modules</u> 1) ODA loan project evaluation methods (lectures and group exercises) - Lecture - Group exercises 2) Other evaluation methods (lectures) - Policy level evaluation - Evaluation systems of international aid organizations <u>II. Evaluation system workshop modules</u> - Presentation on past joint evaluations - Group discussions on issues and measures for strengthening evaluation capacity of participating countries - Preparation of individual action plan for evaluation capacity development
Case Study Materials	Simulation of ex-post evaluation of an irrigation project	Simulation of ex-post evaluation of a transportation (tunnel construction) project

Source: Adopted from the materials for the two seminars, JICA.

The Evaluation Seminar's initial main purpose was to gain participants' understanding of Japan's ODA project evaluation so that they would better receive evaluators from Japan in the event of ex-post evaluation. Over the course of several years the Evaluation Seminar has clearly become a means of ECD. Now the seminar is the first step in JBIC's assistance in ECD in ODA loan projects, aiming to strengthen the evaluation capacity of participating organizations so that they can eventually evaluate their ODA projects by themselves. As the purpose of the Evaluation Seminar shifted, the scope of the seminars have widened far beyond merely explaining Japan's ODA pro-

ject evaluation method. The evaluation seminars now include workshops in which participants can discuss evaluation-related issues among themselves as well as with guest speakers who previously received JBIC's ECD assistance.

(2) Evaluation seminar participants: inviting various organizations

The Evaluation Seminar primarily targets management level officials of ODA planning/ coordinating agencies and executing agencies of ODA loan projects. The actual participants vary and include both working level staff and senior level officials. In addition, several participants come from the executing agencies of technical assistance projects (from Japan as well as from other donor countries). Consequently, every year, the group turns out to be quite a mix of backgrounds and participants are afforded a good opportunity to obtain new ideas.

(3) Evaluation training modules: creating/ identifying demands for learning evaluation methods

Training on loan project evaluation methods form the core of the Evaluation Seminar's evaluation training modules. The Evaluation Seminar's evaluation training modules consists of lectures and simulation exercises on ex-post evaluation of ODA loan projects. Using case studies based on real evaluation cases, participants practice essential steps such as evaluation design, preparation of data collection tools, data analysis and drawing group conclusions. Every year, participants focus on lectures and exercises, and enjoy group discussions, which turn out to be very lively. In the 2006 Evaluation Seminar, pre-departure exercises were introduced for the first time. Individual participants prepare answers to exercise questions before they come to Japan, and during the Seminar sessions, they work in group to discuss their answers and come up with the evaluation conclusions. It was observed that the pre-departure exercises enhanced participants' basic understanding of training contents and their willingness to attend the Evaluation Seminar to check their answers with other participants (Details of Evaluation Training ODA Loan Projects (Example of 2007 Seminar) is provided in Annex).

Japan's ODA loan project evaluation is based on the DAC aid evaluation principles and therefore it has many similarities to evaluation procedures of other donor organizations. This can satisfy demands of participants to learn internationally-established evaluation methods.

Below are some of the comments from participants of past few years:

- Pre-departure exercises were useful in acquiring knowledge of evaluation.
- Group discussions and presentations based on pre-departure exercises were useful.
- All participants actively participated in discussions.
- Step-by-step approach enhanced understanding.
- Learned experiences and opinions of other countries.
- Need more variety in case studies.
- Evaluation trainings should be synchronized to the conduct of evaluation so that agencies related to the target projects can get necessary skills for evaluation.

In 2006, a participant from Egypt found the evaluation training useful for establishing a monitoring and evaluation system, and invited the instructors to Egypt to hold a similar training for a larger number of officers in charge of ODA planning and management. Traveling costs (to the seminar and back) and training costs were shared by the Egyptian and Japanese sides.

(4) Evaluation systems workshop modules: creating/ identifying demands for institutional enhancement on evaluation

Since 2004, evaluation systems workshop modules have been incorporated into the Evaluation Seminar in order to identify the evaluation demands of participating organizations in institutional development, and to encourage an exchange of information and new ideas on how to improve a country's evaluation system. In recent years, the workshop part has included two modules: (i) presentations by guest speakers from developing countries on their experience in ECD conducting joint evaluations (see Section 3); and (ii) discussion on the problems participating countries have in carrying out evaluation and the possible measures participating countries can take to strengthen their evaluation systems.

All past guest speakers for the first module have been involved in joint ex-post evaluations of Japan's ODA loan projects in the past. They include officers of ODA planning/ coordinating agencies or executing agencies, or private consultants from Thailand, Indonesia, Tunisia, the Philippines and India. Some of the guest speakers have been graduates from past Evaluation Seminars, which served as the starting point of their involvement in joint evaluations.

Guest speakers' presentations generally include (i) profiles and evaluation systems of their organizations/ countries, (ii) information on the project

they jointly evaluated with the Japanese side, and (iii) issues and recommendations related to evaluation. Those presentations are given after participants have completed the evaluation training modules and therefore have acquired knowledge and skills of ODA loan project evaluation. In past Seminars, participants showed an interest in the guest speakers' presentations, especially in issues related to project implementation. Following the presentations, participants exchange ideas and engage in lively discussions.

The second module, a workshop on issues related to evaluation systems, takes place either before the evaluation training module or after the guest speaker presentations. The workshop consists of several components such as (i) briefing by each participant on current evaluation system of his/her country, (ii) discussion on problems in the conduct of evaluation, and (iii) group discussion and individual work on measures to strengthen evaluation systems. Guest speakers from the first module sometimes join the group and help participants consider new ideas about possible measures to be taken. At the end of the workshop, each participant presents an action plan for strengthening the evaluation system of his/her country.

We could easily predict a lot of the comments expressed by participants about problems and measures in their evaluation systems. Some common features are as follows:

- Although evaluation is not given a high priority in many countries, officers in charge of ODA planning and coordination at least understand the need to strengthen their country's evaluation system.
- In many countries, although ODA planning/ coordinating agencies are in a position to facilitate or coordinate evaluation, it's the executing agencies that carry-out the substantial work involved with evaluations. Consequently, ODA planning/ coordinating agencies tend to raise issues on how to mainstream evaluation in aid management, whereas officers in charge of ODA planning and coordination raise issues in regards to lack of resources to conduct evaluation.
- Among different stages of monitoring and evaluation, the implementation stage is given a relatively higher priority (though most of the time as part of management activities of on-going projects using funds of the concerned projects). On the other hand, many countries are unable to afford ex-post evaluations, which are conducted after the closure of the project accounts and withdrawal of the donors. Many participants, particularly those from ODA planning/ coordinating agencies, acknowledge the importance of ex-post evaluation, but also face a situation in

which their scarce resources are needed to make their on-going projects successful.

- Therefore, they need external resources particularly for ex-post evaluation and its feedback. In particular they need financial support to hire evaluation consultants and hold evaluation feedback seminars for their citizens.

2-2 Joint Evaluation Program in Partner Country

(1) Overview

The next step of the ECD model for ODA loan project evaluation is the joint evaluation program. The major common purposes of joint evaluation programs are (i) strengthening of evaluation capacity ODA planning/ coordinating agencies, executing agencies and/or operation and maintenance (O&M) agencies and (ii) harmonization of evaluation procedures of both sides. The main activity is the joint ex-post evaluation of ODA loan projects (technical transfer through OJT), but some joint evaluation programs include other types of technical assistance such as evaluation-related training and study.

So far, JBIC has conducted joint ex-post evaluation of ODA loan projects with such countries as Thailand, Indonesia, Tunisia, Sri Lanka, Malaysia, India, the Philippines, and Vietnam. Among these, the program with Indonesia, the Philippines and Vietnam are based on three-year Memorandum of Understandings (MOUs) between the ODA planning/ coordinating agency of each country and JBIC ⁵.

In principle, this type of joint evaluation is different from the “conventional” ex-post evaluation of ODA loan projects carried out by Japanese external evaluators: in joint evaluation, evaluators from partner countries are informants, as in conventional ex-post evaluations, but also make evaluation plans, conduct data collection and analysis and write evaluation reports (Table 3).

In practice, the degree of participation of partner countries differs widely, case by case. Based on our participation as the Japanese counterpart in past joint ex-post evaluation cases with Indonesia in 2004 and Vietnam in 2007 and 2008 ⁶ (as well as being involved to some extent in two joint ex-post evaluation cases in Sri Lanka in 2005 and in the Philippines in 2007), we have noticed a sharp contrast between Indonesia and Vietnam in that the Vietnam

⁵ MOUs with Indonesia and the Philippines were signed in 2005, and the MOU with Vietnam was signed in 2007.

⁶ Excluding the joint evaluation program funded by the Ministry of Foreign Affairs in 2005 (See Section 3).

Table 3. Roles of concerned agencies in conventional and joint evaluations

● : leading role ○ : provision of comments

ID: Indonesia (2004) VN: Vietnam (2007 & 2008)

Work item	Conventional evaluation by Japanese side				Joint evaluation between Partner and Japanese sides			
	Partner country		Japan		Partner country		Japan	
	Execut- ing, O&M agencies	ODA Planning agency	External evaluator	JBIC	Execut- ing, O&M agencies	ODA Planning agency	External evaluator	JBIC
Preparation of TOR				●		○ ID ● VN		● ID ● VN
Evaluation planning/ design			●	○	● VN	○ ID ● VN	● ID ● VN	○ ID ○ VN
Provision of information	●	●		●	● ID ● VN	● VN		● ID ● VN
Collection and analysis of information			●		● VN	● VN	● ID ● VN	
Report writing	○	○	●	○	○ ID ● VN	○ ID ● VN	● ID ● VN	○ ID ○ VN
Feedback of evalua- tion results			●	●	● VN	● VN	● ID ● VN	● ID ● VN

Source: Prepared by the author.

evaluation was more “joint” than the Indonesian one. In Section 3 we discuss our experience in the joint evaluation programs in Vietnam in detail, focusing on the factors for its success.

(2) Joint evaluation with Indonesia in 2004: findings from a low participation case

Why was the Indonesian counterpart’s participation in the joint evaluation low? During the implementation of the joint evaluation program in 2004⁷, we observed several conditions as follows:

On the side of the executing agency:

- There was little interest in evaluation of completed projects at the exe-

⁷ The target project of the joint evaluation with Indonesia in 2004 was the Jakarta Fishing Port/ Market Development Project (4). The evaluators were Haraguchi and representatives of the Ministry of Marine Affairs and Fisheries, the executing agency of the project. The facilitators of the evaluation were former JBIC officers and the Directorate of Monitoring and Evaluation, National Development Planning Agency (BAPPENAS). For the evaluation results, see the Ex-post evaluation report, Jakarta Fishing Port/ Market Development Project (4), *Evaluation Highlights on ODA Loan Projects 2005*, JICA (http://www.jica.go.jp/english/operations/evaluation/jbic_archive/post/2005/pdf/2-04_full.pdf).

cuting agency. When the joint evaluation was planned, the ODA planning/ coordinating agency's role was that of the evaluator, and the executing agency's role was that of the informant or the "evaluatee." However, when the role of the ODA planning/ coordinating agency was clarified to be the facilitator of evaluation, the executing agency suddenly became the evaluator without knowing what that meant. Although the representative from the ODA planning/ coordinating agency had participated in the ODA Loan Project Evaluation Seminar in Japan, the representative from the executing agency had not.

On the side of the ODA planning/ coordinating agency:

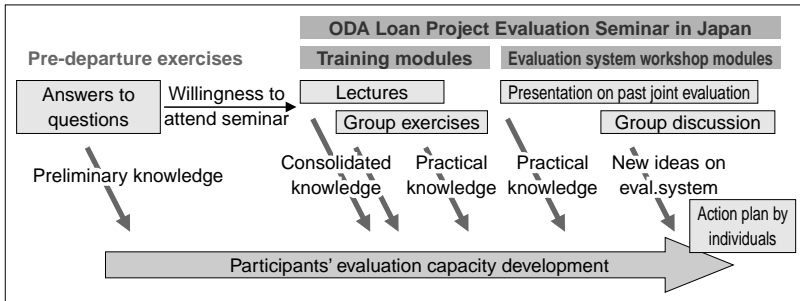
- The ODA planning/ coordinating agency, which had initially requested the former JBIC for the joint evaluation and attended the ODA Loan Project Evaluation Seminar, wanted to outsource the evaluation to local consultants rather than directly undertake evaluation work such as data collection and report-writing. Therefore, as JBIC hires Japanese consultants as the Japanese-side evaluator, the Indonesian ODA planning/ coordinating agency's main request was to hire local consultants as the evaluator of the Indonesian side.
- We confirmed that the officers in charge at the ODA planning/ coordinating agency had a good knowledge of evaluation as well as willingness to promote evaluation as a means to improve project management by attending future JBIC ODA Loan Project Evaluation Seminars and other meetings/ workshops. However, the organization lacked the human and financial resources to put them into practice.

On both sides of executing agency and ODA planning/ coordinating agency as well as JBIC

- The arrangements among participating organizations for communication and coordination were not clearly defined. As a result, key persons of the executing agency often failed to attend evaluation meetings and field trips, and executing agencies' interest in evaluation did not manifest itself until the end of the process.

2-3. Factors for a successful ECD model

Based on our experience we have tried to identify some factors for effective ECD assistance as below;

Figure 1: Factors of Learning through ODA Loan Project Evaluation Seminars

Source: Prepared by the author.

(1) Assistance in institutional development

As already mentioned, participants in the Evaluation Seminars showed a high level of interest in learning evaluation methods based on international standards. Despite this interest, in the discussions on how to improve country evaluation systems, participants seldom raised methodology-related topics, focusing instead on institutional, legal and financial matters. This suggests that although knowledge and skills of evaluation methods are important, ECD should include institutional development as well so that officers can put the knowledge they acquired into practice in their organizations/ countries.

In this respect, in addition to the transfer of evaluation methods, evaluation trainings and seminars could better serve ECD by providing participants with opportunities to acquire various ideas on how to realize institutional development. JBIC does incorporate aspects of institutional development into its ODA Loan Project Evaluation Seminars, through guest speaker presentations in which they present their experiences in and the merits of practicing the learned evaluation methods, and through workshops in which participants with different background discuss evaluation systems.

(2) Raising executing agencies' awareness of evaluation

One of the common observations from the above-mentioned cases of ODA Loan Project Evaluation Seminars and the joint evaluation program in Indonesia is that, although in many countries executing agencies are the ones who actually conduct project evaluation, executing agencies, tend to focus on the approval for and implementation of the project, and therefore they might have less interest in evaluation throughout the project cycle including ex-post evaluation and its feedback.

To raise executing agencies' awareness of evaluation, the donor side

could encourage them to participate in evaluation so that they accumulate evaluation practices in their organizations. Also, capacity of ODA planning/ coordinating agencies to facilitate evaluation, i.e., encouraging executing agencies to carry out evaluation and continuously sending them messages about the importance of evaluation, should be enhanced as well. In all this, division of responsibility of different organizations in evaluation should be clearly identified and shared among all concerned parties.

(3) Transfer of knowledge and skills of monitoring and mid-term evaluation

So far, ODA Loan Project Evaluation Seminars have primarily dealt with methods of ex-post evaluation (conducted after two years of project completion). However, many seminar participants give higher priorities to monitoring and evaluation during project implementation. JBIC could provide its expertise in the methods for monitoring and evaluation during project implementation as it promotes a consistent monitoring and evaluation system throughout the project cycle (i.e. from ex-ante to ex-post).

Of course, all the different stages of evaluation share some common features, and ex-post evaluation is a good subject to cover all aspects of evaluation. To further enhance participants' satisfaction with future Evaluation Seminars, such seminar could, for example, add concrete know-how of implementation monitoring and mid-term evaluation in addition to covering common features of evaluation. Executing agencies in particular would welcome this addition since they are mostly interested in the implementation phase.

(4) Connecting evaluation seminars to further assistance in ECD (country evaluation seminars and joint evaluation programs)

Since the first year, satisfaction of participants with the Seminars has been quite high, and some participating organizations have entered the next step of JBIC's ECD assistance. The major path is toward a technical cooperation program between JBIC and a specific developing country, whose main component is joint evaluation of ODA loan projects (see next sections for more details). Besides technical cooperation programs, there was a case where the participants invited the instructors to his country to hold a similar evaluation seminar for more officials concerned. The government of Egypt and JBIC shared the cost for the seminar in Egypt.

(5) Demand for outsourcing of evaluation and incorporate assistance in evaluation management

A frequently discussed issue among developing countries is their lack of human and financial resources and naturally there is a demand for financial assistance to hire local evaluators. Providing such financial assistance, however, could lead to a hollowing out of evaluation know-how in the counterpart organizations if they overly rely on such hired consultants.

Outsourcing itself might work well in many developing countries, especially where government evaluation resources are scarce. However, care should be taken so that the concerned government agencies acquire sufficient capabilities to manage the evaluation process and utilize results to fulfill their responsibility in the evaluation, i.e. evaluation management including the facilitation of the evaluation and quality control of the work carried out by external evaluators.

3. Case Study of Joint Evaluation in Vietnam

This section, covers two cases of joint evaluation programs in Vietnam. The first case is the joint *program* evaluation between the Ministry of Planning and Investment (MPI) of Vietnam and the Ministry of Foreign Affairs (MOFA) of Japan in 2005, and the second case is the first-year program of the three-year technical cooperation on joint evaluation between MPI and JBIC in 2007. It could be said that the Vietnam side more actively participated in the 2007 joint evaluation and therefore it was more effective for ECD than the 2005 evaluation. First, we will present the basic profiles of the two cases, and then compare their performance and outcomes to identify the key factors that influenced the differences.

3-1 Background and Outline of Two Joint Evaluations in 2005 and 2007

3-1-1 MPI-MOFA Joint Evaluation in 2005

(1) Background

Vietnamese representatives from MPI first proposed the idea of a MPI-MOFA joint evaluation at the “Third ODA Evaluation Workshop ⁸” in November 2003 hosted by MOFA. The purpose of the proposal was to conduct a joint monitoring and evaluation exercise with a possible impact on capacity development in Vietnam. In July 2005, MPI and MOFA agreed to

execute joint evaluation activity on a Japanese ODA program for transport sector development in the Red River Delta area⁹. They also agreed to adopt MOFA's "ODA Evaluation Guidelines" as the basic evaluation method for the joint evaluation study.

The objectives of the Joint Program Evaluation Study were: (i) to plan and execute a joint program evaluation study of the Japanese ODA program for the transport sector development in the Red River Delta area, and (ii) to promote Vietnamese understanding of program evaluation on ODA through the participatory approach to the study.

(2) Outline

Evaluation Method

The MPI-MOFA joint evaluation in 2005 used the ODA evaluation practice established by MOFA and stated in "*ODA Evaluation Guideline* (the first version, March 2003).¹⁰" According to the *Guideline*, this joint evaluation was classified as a "Program-level Evaluation," and in particular it was further classified as a "Sector Program Evaluation." The *Guideline* adopts a comprehensive evaluation method for the Program-level Evaluation (Sector Program Evaluation), in which the object is evaluated from three points: purpose, process, and results.

Object of Evaluation

Since a comprehensive sector program covering all Japanese ODA projects in the transport sector of Vietnam did not exist, a "quasi-program," (based on one conducted by JICA "the Master Plan Study of Transport Development in the Northern Part in the Socialist Republic of Vietnam (1994)"¹¹) was expediently developed exclusively for the evaluation purpose. The "quasi-program"

8 In response to growing awareness of the importance of donor-partner cooperation in tackling development challenges and global development issues, MOFA has hosted the "ODA Evaluation Workshop" regularly since 2001 inviting representatives of 18 Asian partner countries together with bilateral and multilateral development agencies and banks including the World Bank, the Asian Development Bank (ADB), the United Nations Development Program (UNDP), the Organization for Economic Cooperation and Development (OECD), the Japan International Cooperation Agency (JICA), and the former Japan Bank for International Cooperation (JBIC). The first workshop was held in November 2001 (in Tokyo) followed by the second workshop in November 2002 (in Tokyo), the third workshop in November 2003 (in Tokyo), the fourth workshop in January in 2005 (in Bangkok), the fifth workshop in January 2006 (in Tokyo) and the sixth workshop in November 2007 (in Kuala Lumpur).

9 The final report is available at MOFA's web site at "<http://www.mofa.go.jp/policy/oda/evaluation/2005/vietnam.pdf#search='MOFA joint evaluation Vietnam'>".

10 The MOFA has regularly updated the *ODA Evaluation Guideline* and the latest version (the fourth version) was published in May 2008.

was given the name, “The Japanese ODA Program for transport infrastructure development in the Red River Delta area.”

The Red River Delta Transport Development Program comprised a group of Japanese ODA projects carried out during the 1994 to 2004 target period including, 13 ODA loan projects, 2 grant aid projects, 2 technical cooperation projects, and 8 development studies (the list of Japanese ODA projects under the program is provided in Annex Table 2-1).

Process of Joint Evaluation Activities

The MPI-MOFA joint evaluation activities were conducted in the following three stages; (i) evaluation planning, (ii) data collection and analysis, and (iii) conclusion of evaluation. The overall implementation period of the MPI-MOFA joint evaluation was about six months from July 2005 to February 2006. Since this joint activity was aimed at developing the Vietnamese capacity on program evaluation, the process employed an On the Job Training (OJT) style in the workshop, training, and collaborative activities. However, feedback of the evaluation results was not necessarily included in the framework of MPI-MOFA joint evaluation.

(i) Evaluation Planning Stage (from July to August 2005)

Both evaluation teams (Vietnam and Japan) prepared and mutually agreed to the objective framework of the Red River Delta Transport Development Program and the “Evaluation Framework.” An “ODA Evaluation Seminar” was also held to gain mutual understanding of issues including; the purpose of the study, the proposed evaluation methodology, the research plan, and the implementation schedule.

(ii) Data collection and analysis stage (from September to November 2005)

Actual research activities and data collection were then conducted. For the data collection, a combination of multiple survey methods were employed including the questionnaire survey with supplementary interview, interview survey, beneficiary survey with semi-structured interviews, direct observa-

11 The Master Plan 1994 was the first master plan to target the transport sector in the northern part of Vietnam, proposing a complex integrated network of transport systems and services in the four sub-sectors including the road, railway, sea and port, and inland waterway transport sectors. This evaluation study utilized the framework of the Master Plan 1994 in order to create a “quasi-program” as an object for the study.

tion by project site visit, and document review. The Japanese team did the main work on data compilation and analysis, and also produced the preliminary evaluation results.

(iii) Conclusion of evaluation stage (from December 2005 to February 2006)

The Japanese and Vietnamese teams discussed the preliminary evaluation results and carried out a revision of the results necessitated by the critical comments from the Vietnamese team.

The draft report was circulated to the related ministries and agencies in both Japan and Vietnam for their review and comments. Based on these comments, the final draft report was produced.

Formation of Joint Evaluation Team

On behalf of the government of Vietnam, seven officers from MPI and MOT participated in the joint evaluation. In addition, three experts from the Transport Development Strategy Institute (TDSI) and VAMESP II ¹² also joined the Vietnam core-team members as observers.

The MOFA evaluation team included four officers from MOFA including the Embassy of Japan in Hanoi and five consultants (two Japanese evaluation experts and three national consultants). In addition, five research assistants were temporarily employed to support the economic impact study.

Cost of Joint Evaluation

MOFA bore the employment and traveling cost for the Japanese evaluation experts, the national consultants, and the research assistant as well as the costs for the workshop, training and printing materials. MOFA also provided the transportation cost for the field survey for the Vietnamese core-team. VAMESP II funded the travel expenses for the field survey for the Vietnamese core-team members. Cost sharing by the GOV was not possible because the Evaluation Cost Norms had not been prepared yet.

The outline of the MPI-MOFA joint evaluation above is summarized in Annex Table 2-2.

¹² The Vietnam Australia Monitoring and Evaluation Strengthening Project Phase II (VAMESP II), which is a technical cooperation project funded by AusAID.

3-1-2 MPI-JBIC Joint Evaluation in 2007

(1) Background

In response to a series of international and national consensus building events since the middle of the 2000s, such as the Millennium Development Goals (MDGs), the Vietnam Development Goals (VDGs), the Paris Declaration on Aid Effectiveness (March 2005) and the Hanoi Core Statement (September 2005), Vietnam has made progress on improving ODA management. The Government of Vietnam has become keen to improve its capabilities in the management of ODA programs and projects, and the GOV set out the ODA Strategic Framework in 2006-2010 and issued the Regulation on Management and Utilization of ODA (Decree No. 131/2006/ND-CP) in 2006.

Decree No. 131 clearly defined the roles and responsibilities of different organizations/ agencies engaged in ODA M&E activities. In order to fulfill such responsibilities, the MPI developed the Framework for Monitoring and Evaluation of ODA Programs and Projects in 2006-2010 Period (Decree No. 1248/2007/QD-BKH) and the Action Plan for the Framework.

At the same time, the GOV has worked on M&E capacity building such as the development of the Monitoring and Evaluation Manuals, training/ seminars and pilot evaluations, many of which have been a part of VAMESP I & II.

With the mutual interest in further strengthening the GOV's evaluation capacity and in establishing an effective joint evaluation scheme, JBIC and MPI concluded a Memorandum of Understanding (MOU) in ECD in July 2007 for collaboration on joint evaluation.

The objective of the MOU is to join efforts and maintain an ongoing working relationship to achieve; (i) effective and efficient implementation of JBIC assisted ODA projects through improvement of evaluation and feedback mechanism of evaluation results to the implementers and policymakers of the GOV; and (ii) institutional improvement through harmonization of evaluation mechanisms of the GOV and JBIC.

The Joint Evaluation Program 2007 was the first-year Implementation Program of Joint Evaluation based on the MOU. The transportation sector was selected as the target sector of the Program.

(2) Outline

Evaluation Method

MPI-JBIC joint evaluation used the *Monitoring and Evaluation Manual*:

*Evaluation Practice Module*¹³ published by MPI in May 2007. VAMESP II developed methods for the GOV's ODA-M&E system referencing the evaluation methods provided in the Project Cycle Management (PCM) handbooks¹⁴ of Japan. In addition to methods from Japan, VAMESP II frequently referred to the evaluation methods of IFAD.¹⁵ The type of evaluation was an ex-post project evaluation.¹⁶

Object of Evaluation

MPI-JBIC selected three ODA loan projects in the transport sector for the target projects of joint evaluation as listed below:

- (i) National Highway No. 5 Improvement Project (1) (2) (3)
- (ii) National Highway No. 1 Bridge Rehabilitation Project (I-1) (I-2) (I-3) (II-1) (II-2) (II-3)
- (iii) Hanoi-Ho Chi Minh City Railway Bridge Rehabilitation Project (1) (2) (3)

Process of Joint Evaluation Activities

The MPI-JBIC joint evaluation conducted activities as per the following four stages (i) evaluation planning, (ii) data collection and analysis, (iii) conclusion of evaluation and (iv) feedback. The overall implementation period of the MPI-JBIC joint evaluation was about eleven months during the period between August 2007 and June 2008, but the major joint activities were carried out from November 2007 to June 2008.

Similarly to the MPI-MOFA joint evaluation in 2005, the 2007 MPI-JBIC joint evaluation employed an OJT style through workshops, training, and collaborative activities in order to strengthening GOV's evaluation capacity.

Since the joint evaluation is a tool for evaluation capacity development of the Vietnamese counterparts as well as promotion of Vietnamese ownership, the individual activity in each stage was designed by the participatory

13 The *Manual* consists of the Monitoring Practice Module and the Evaluation Practice Module, and covers all essential matters such as purposes of M&E, M&E criteria, M&E planning, data collection and analysis, reporting, case studies of VAMESP II pilot M&E activities.

14 The PCM handbooks were developed by FASID, Japan. The PCM methods in Japan were developed based on several guidelines using Logical Framework Approach including those from IFAD, the ODA-M&E methods of GOV and Japan share a number of common aspects.

15 In developing the M&E Manual, VAMESP II conducted a number of pilot evaluations for ODA program/projects. Various methods were tested and improved in those pilot evaluations.

16 In the JBIC's evaluation system, three types of project evaluation such as ex-ante evaluation, mid-term review, and ex-post evaluation were to be conducted. The target projects for ex-post evaluation must be the projects after two years of its completion.

approach and the works were shared based on the consensus building between the Vietnamese and JBIC team members. Principally the JBIC team took leading roles for preparation of the evaluation framework, workshops and training, management of field works and report writing, whereas the Vietnamese core-team members shared in most of the practical activities.

(i) Evaluation planning stage (from September to November 2007)

The JBIC team drafted the implementation plan (IP2007) and held a kick-off meeting as well as individual meetings from the 25th to the 28th of September 2007 with MPI, MOT, Vietnam Railways, Project Management Unit No. 5 (PMU5), Project Management Unit No. 18 (PMU18), Railway Project Management Unit (RPMU) and VAMESP II. In these meetings, related parties discussed the overall framework for IP2007 including its methodology, implementation schedule, and implementation structure.

Based on the above discussion and agreement, the JBIC team prepared the draft Evaluation Planning Framework harmonized with the MPI's *Monitoring and Evaluation Manual*.

The joint evaluation team comprising the Vietnamese core-team and JBIC team members was officially established during the evaluation workshop on the 1st and the 2nd of November 2007. Then the joint evaluation team intensively worked on preparing questionnaires, finalizing the evaluation planning framework and tools, scheduling field surveys as well as implementing the pilot beneficiary survey, including interviewing private companies in the industrial park and holding focus group meetings.

(ii) Data collection and analysis stage (from December 2007 to January 2008)

The Vietnamese core-team conducted the data collection survey according to the field survey schedule. Vietnamese and Japanese team members jointly conducted every data collection activity. They employed a combination of the multiple survey method which included the questionnaire survey with supplementary interview, the interview survey, the beneficiary survey with semi-structured interview, focus groups, direct observation through project site visits, and document review.¹⁷

After the data collection stage, each working group worked on the analysis of the collected data and follow-up data collection. Work was allocated to each working group for preparation of the evaluation summary. Whilst the Vietnamese core-team was responsible for data analysis and drafting evalua-

tion summary for relevance, efficiency, and effectiveness of the projects, the JBIC team focused on evaluation of impact and sustainability of the projects. This division of responsibility was decided based on the following two considerations. First, it was relatively easy for the PMUs, the “core” of the Vietnamese core-team, to work on the matters directly related to their organizational mandate, i.e., implementation of the project. Second, findings on the matters related to project implementation would be more useful for the PMUs than external evaluators. Evaluation of the project in relation to high-level development objectives and sustainability after the project completion, on the other hand should require external viewpoints.

(iii) Conclusion of evaluation stage (from February to May 2008)

Both side shared, criticized, revised, and then consolidated the evaluation analysis results into an evaluation report. Each working group presented their evaluation summary for each project at the internal workshop on the 29th of February 2008. Based on the results of the workshop, the JBIC team drafted full evaluation reports and circulated the drafts around to the concerned Vietnamese ministries and agencies as well as to JBIC. The JBIC team then finalized the draft evaluation reports.

(iv) Feedback stage (June 2008)

In order to share the results of the evaluation results of the three target projects, a feedback workshop was conducted on the 23rd of June 2008. Representatives of concerned ministries and agencies of Vietnamese government, members of the mass media as well as major foreign donors were invited.

Formation of Joint Evaluation Team

The Vietnamese core-team was comprised of 22 officers from MPI, MOT, PMU18, PMU5, the Vietnam Railways Corporation, the Railway PMU, the

17 NH-5 working group conducted: 5 focus groups (FGs) in Hanoi, Hai Duong, Hung Yen provinces with participation of 94 local residents, semi-structured interviews (SSIs) to the people’s committees of Hung Yen Province, Hai Duong Province and Hai Phong City, and SSI to 9 companies in Hung Yen, Hai Duong and Hai Phong provinces.

NH-1 and railway project working groups conducted: 7 FGs in Quang Binh, Quang Nam, Khanh Hoa, Bac Giang provinces with participation of 118 local residents, SSI to the people’s committees of Quang Nam, Khanh Hoa, Binh Dinh, Binh Thuan, Ho Chi Minh City, and Bac Giang provinces, SSI to 13 companies, organizations and transporters in Da Nang City, Khanh Hoa, Binh Dinh, Binh Thuan, Ho Chi Minh City, and Bac Giang provinces, and SSI to 39 train passengers in 4 railway stations including Hue station, Da Nang station, Saigon station and Hanoi station.

Vietnam Road Administration, the National Transport Safety Committee, and the Transport Development Strategy Institute (TDSI).

The JBIC evaluation team was comprised of two Japanese external evaluators (Japanese evaluation consultants¹⁸) and two national consultants. During the field survey, six research assistants were temporarily employed to support the Focus Group and interview survey of the beneficiaries.

Cost of Joint Evaluation

JBIC bore the employment cost for the Japanese external evaluators, the national consultants, and the research assistants (and their traveling costs), as well as the costs for workshops, training and printing materials. JBIC also covered part of the Vietnamese core-team's transportation cost for the field survey. Each Vietnamese ministry and agency bore the travel expenses of field survey for their own core-team members but VAMESP II funded the travel expenses for the staff of MPI and MOT. Cost sharing by the GOV was not possible because the Evaluation Cost Norms had not been prepared yet.

Annex Table 2-2 summarizes the outline of the MPI-JBIC joint evaluation as explained above.

3-2 Comparison of Performance and Outcomes of the Two Joint Evaluations in 2005 and 2007 and Key Factors that influenced the Differences

3-2-1 Comparison of Performance and Outcomes of the Two Joint Evaluations in 2005 and 2007

The degree of commitment and ownership of the Vietnamese core-team was higher in the 2007 MPI-JBIC joint evaluation than in the 2005 MPI-MOFA joint program evaluation. Basically the two joint evaluations took similar steps in the process from the evaluation planning to the conclusion of evaluation. But the Vietnamese participants in 2007 took active roles in all of the processes and substantially contributed to the evaluation activities such as participating in the workshop and training, drafting the evaluation framework, field survey planning, data collection and analysis, conclusion of report and feedback stages. The Vietnamese core-team members in 2007 took responsibility for tasks agreed to in the joint evaluation team. For example, they carried out a part of the interview survey to the informant agencies and

18 The authors of this chapter.

beneficiaries without the assistance of Japanese team. They worked on the assigned data compilation and analysis regarding relevance, efficiency, and effectiveness criteria and drafted the evaluation report for their assigned part. At the feedback seminar, they performed an excellent presentation on the evaluation results of each project. Their motivation to learn was high. In particular the participants from the executing agencies and PMUs of the projects who were directly involved in implementation, operation and maintenance of the projects were very enthusiastic about the joint evaluation activities. In fact, JBIC team and the Vietnam core-team shared many of the tasks and activities according to the agreed upon division of tasks for each stage.

At the final stage of the 2007 MPI-JBIC joint evaluation, the JBIC team conducted a rapid questionnaire survey¹⁹ for the Vietnamese core-team members to ascertain useful lessons and recommendations for future joint evaluation programs. According to the collected answer from 14 participants out of 23, all of 14 respondents perceived that they learned from the joint evaluation.²⁰ Similarly, all of 14 respondents answered that the experience of the joint evaluation will be useful for their business.

According to responses to an open-ended question, members replied that they learned both (i) technical matters such as overview of evaluation, evaluation planning, data collection and analysis, report-making, etc., and (ii) matters more related to management such as importance of evaluation in investment project, how to utilize evaluations in project management, teamwork, lesson-learning for future projects, etc.

In fact, the joint evaluation provided opportunities for the Vietnamese participants to review the whole process of the project from the planning to the implementation and to confirm objectively whether their projects produced the expected outcomes and impacts.

Many of the Vietnamese participants appreciated that they could directly listen to the project beneficiaries including the local peoples as well as the project related agencies and authorities and they learned from the people. Based on such experiences, they proposed practical and constructive recommendations for future projects. The joint evaluation activities stimulated the awareness of the Vietnamese participants on the importance of evaluation in the project cycle management.

19 The questionnaire was sent to 21 members and 2 assistants, and a total of 14 responses were collected (2 from MPI, 2 from MOT, 8 from PMUs, 1 from VRA and 1 from TDSI). The question was "Did you learn new things through the Joint Evaluation Program 2007?"

20 Among 14 responses, 4 respondents answered "Yes, very much" and 10 respondents answered "Yes".

Therefore, it can be concluded that the ECD thorough the 2007 MPI-JBIC joint evaluation was quite successful.

Regarding the 2005 MPI-MOFA joint evaluation, the degree of commitment and ownership was moderate and the attitude of the participants was relatively passive. They participated in the workshop and training, field survey for data collection, and data analysis process, but the Japanese team always took the lead in evaluation planning, data collection and analysis, drawing evaluation conclusions and report writing. For instance, they participated to the interview survey and asked some questions to the informant ministries and agencies, but did not take part in the data compilation and analysis. Instead of taking part in data compilation and analysis, the Vietnamese side took responsibility for providing comments on the evaluation summary results and the final reports prepared by the Japanese team.

Another aspect is a lack of evaluation feedback in the 2005 MPI-MOFA joint evaluation. Because the 2005 joint evaluation was intended as a pilot evaluation to exercise the technical transfer of ODA evaluation from Japan to Vietnam emphasizing aspects of evaluation method and practice, feedback activities were initially not given much consideration in designing the framework of the MPI-MOFA joint evaluation. But the results of the MPI-MOFA joint evaluation were published in MPI's *Monitoring and Evaluation Manual: Evaluation Practice Module* as a case study of program evaluation with the partnership of the foreign donor.

It can be concluded that the ownership, motivation and performance of the Vietnamese core-team members in 2007 was higher than that in 2005, hence the ECD assistance was much more effective at the MPI-JBIC joint evaluation in 2007. The possible key factors which influenced these differences might be (i) the development of a legal framework for M&E in Vietnam, (ii) the proceeding M&E projects with partnership of donors, (iii) the type of evaluation and methodology adopted, (iv) the inputs from Japanese side, and (v) the continuity of the person in charge.

3-2-2 Key Factors for the Differences between the Two Joint Evaluations

(1) Development of a legal framework for M&E in Vietnam

Firstly, the readiness of the Vietnamese government to think about ODA evaluation was much higher in the mid 2000's. In response to the Paris Declaration on Aid Effectiveness²¹ in February 2005 and the Hanoi Core Statement²² in September 2005, the GOV institutionalized and perfected the

Table 4. Key Milestones for the Development of Legal Framework regarding M&E of ODA in Vietnam

Year/month	Under Law Legal Document
June 2003	Decree No. 61/2003/ND-CP of the Government on Prescribing the Functions, Tasks, Powers and Organizational Structure of the Ministry of Planning and Investment
February 2005	Paris Declaration on Aid Effectiveness
September 2005	Hanoi Core Statement
June 2006	Decision No. 150/2006/QĐ-TTg of the Prime Minister on Issuance of the Action Plan of the Government to implement the National Strategy on Borrowing and Paying External Debts up to the year 2010
November 2006	Decree No. 131/2006/ND-CP of the Government on Issuance of the Regulation on Management and Utilization of ODA
December 2006	Decision No.290/2006/QĐ-TTg of the Prime Minister on Issuance of the Strategic Framework for Official Development Assistance Mobilization and Utilization for the period 2006-2010
March 2007	Circular No.03/2007/TT-BKH of MPI on Guiding the Organizational Structure, Functions and Responsibilities of ODA Program or Project Management Units
June 2007	Decision No. 94/2007/QĐ-TTg of the Prime Minister on Approval of Action Plan for Implementation of the Strategic Framework for Official Development Assistance Mobilization and Utilization for the period 2006-2010
July 2007	Circular No.04/2007/TT-BKH of MPI on Guiding on the Implementation of Management and Utilization of ODA
October 2007	Decision No. 1248/2007/QĐ-BKH of MPI on Issuance of the Framework for Monitoring and Evaluation of ODA Programs and Projects in 2006-2010 Period together with the Action Plan on Establishment and Operation of the National System on Monitoring and Evaluation of ODA Programs and Projects in 2006-2010 Period

Source: MPI

monitoring and evaluation of ODA programs and projects through issuing various decrees, decisions, and circulars on ODA management. Table 4 lists the key milestones for the development of a legal framework regarding M&E of ODA in Vietnam.

Decree 131 states that evaluation of ODA programs and projects in

- 21 The Paris Declaration on Aid Effectiveness in 2005 sets out five principles to improve aid effectiveness, (i) Ownership, (ii) Alignment, (iii) Harmonization and Simplification, (iv) Management for Results, and (v) Mutual Accountability.
- 22 The main points of the Hanoi Core Statement (HCS) are: (a) the Vietnamese Government and donors agreed to carry out strategic and monitorable activities to realize commitments stated in the Paris Declaration, taking into account specific conditions of Vietnam into local commitments; (b) in order to raise awareness and changing behavior for better aid effectiveness to support development objectives, partnership commitments include 28 separated commitments and common commitments were produced, (c) targets to be achieved by 2010 and their 14 indicators were set up. These indicators are the basis for monitoring and evaluation for implementation of the Hanoi Core Statement.

Table 5. Outline of Decision No. 1248/2007/QĐ-BKH of MPI

Strategic Orientation of Framework for Monitoring and Evaluation of ODA Programs and Projects 2006-2010 and Action Plan for Implementation of the Framework (MPI Decision No. 1248/2007)	
Objective 1	Develop a unified information system to ensure the operation of national system of monitoring and evaluation of ODA programs and projects (ODA-M&E) <ul style="list-style-type: none"> • Institutionalization of ODA-M&E, development of formats to collect data and information, development of ODA-M&E portal, etc.
Objective 2	Selection and adoption of advanced methodologies and tools in ODA-M&E in conformity with Vietnamese situation <ul style="list-style-type: none"> • Development of manuals, development of evaluation formats and rating methods, computer software, etc.
Objective 3	Professionalization of staff working on ODA-M&E <ul style="list-style-type: none"> • Trainings, establishment of Evaluation Club, establishment of Evaluation Association, etc.
Objective 4	Ensuring the budget for ODA-M&E <ul style="list-style-type: none"> • Establishment of Cost Norms for M&E, financial plans for impact evaluation, etc.
Objective 5	Cooperation with donors on ODA-M&E ²³ <ul style="list-style-type: none"> • Establishment of cooperation mechanism, impact evaluations in respective sectors of transportation, power, health, education and training, rural development and poverty reduction, effectiveness evaluation for the Socio-Economic Development Plan, etc.
Objective 6:	Using the results of M&E in Management for Development Results (MfDR) <ul style="list-style-type: none"> • Publishing monitoring reports, assessing the operation of ODA-M&E
Objective 7:	Integrating and using tools, skills and experiences in the system of ODA-M&E to develop the system of M&E of public investment <ul style="list-style-type: none"> • Recommendation for development of M&E system of public investment based on the assessment of operation of ODA-M&E

Source: Adopted from *Framework for Monitoring and Evaluation of ODA Programs and Projects in 2006-2010 Period*, MPI, 2007.

Vietnam can be carried out periodically or on ad-hoc basis. Periodic evaluation can be conducted at four key stages of the program/ project cycle: (a) soon after commencement (initial evaluation), (b) in the middle of program/ project implementation (mid-term evaluation), (c) upon completion (terminal evaluation) and (d) after the project completion (impact evaluation).

Decree 131 also defines the roles and responsibilities of the government agencies involved in the implementation of ODA programs/projects as follows: (i) PMU: carry out periodic monitoring and evaluation, (ii) Project Owner: guide, urge and support M&E activities performed by PMU, (iii) Line Agency: monitor the progress of programs/projects under its authority and

²³ The text of the *Framework* reads “cooperation with donors,” while the same item in the *Action Plan* reads “impact evaluation”.

review/feedback evaluation results, (iv) MPI: coordinate with other ODA-related agencies, and (v) General Statistical Office (GSO): develop statistical information related to receiving and utilizing ODA.

In order to fulfill such responsibilities, the MPI developed the Framework for Monitoring and Evaluation of ODA Programs and Projects in 2006-2010 Period (Decree No. 1248/2007/QĐ-BKH) and the Action Plan for the Framework. Decision 1248 of MPI sets the strategic orientation for the framework for M&E of ODA Programs and Projects 2006-2010 and action plan for implementation of the framework. Decision 1248 requires (i) selection and adoption of advanced methodologies and tools in ODA-M&E in conformity with Vietnamese situation, (ii) professionalization of staff working on ODA-M&E, (iii) cooperation with donors on ODA-M&E and so on (see Table 5).

At the time of the 2005 MPI-MOFA joint evaluation, the process of the above legal framework for M&E in Vietnam was still in an immature stage. However, by the time of the 2007 MPI-JBIC joint evaluation, the legal environment had undergone much progress.

Institutionalization of M&E by means of Decree No. 131 and other legal instruments made the Vietnamese core-team members in 2007 think that evaluation was their job. Therefore, they were very keen in learning project evaluation methods and practices through joint evaluation activities. The legal environment for M&E in Vietnam motivated the Vietnamese people and their ownership.

(2) Proceeding M&E projects with partnership of donors

Secondly, the GOV has promoted the partnership with donors in M&E of ODA programs and projects. Above all a technical assistance project supported by AusAID – the Vietnam Australia Monitoring and Evaluation Strengthening Project Phase I & II (VAMESP I&II: 2003-2008 ²⁴) made substantial progress developing Vietnam's M&E system.

In the framework of VAMESP, the GOV implemented pilot-based monitoring and evaluation of ODA programs and projects in six line ministries and seven provinces and centrally-run cities. VAMESP has been very successfully

24 VAMESP has three major targets: (i) to develop the monitoring and evaluation methodologies and techniques suitable for national adoption in Vietnam; (ii) to obtain experience and necessary lessons from pilot program and project evaluation for scaling up towards a full system, and (iii) to establish sustainable and professional national monitoring and evaluation through human resource development for ODA related agencies.

Table 6: Major Outputs of VAMESP II

	Major Outputs related to Evaluation Capacity Development
M&E Training	522 participants received Basic training, 221 Intermediate, 42 Advanced.
Basic IT Training	96 participants trained.
On-the-job training	More than 763 participants.
Pilot evaluations	17 completed (including MPI-MOFA joint evaluation in 2005 and MPI-JBIC joint evaluation in 2007).
Competency testing in M&E	More than 200 participants annually tested.
Technical working group meetings	20 meetings attended by 492 participants.
Workshops	11 large workshops attended by total 3,849 participants.
Study tours	5 study tours completed to Thailand, Japan, Malaysia, Australia and Singapore.
Regional networking	Liaising with regional evaluation societies.

Source: VAMESP II Collated Outputs, December 2006 and additional information by JBIC team.

implementing and producing valuable outputs such as: (a) Aligned monitoring format and tools; (b) M&E capacity development for staff of line agencies, PMUs, and provinces/cities (adult learning, on-the-job approach, modular training program at 3 levels); (c) 17 pilot evaluations (including 3 joint evaluations - one with Government of Japan (MOFA), one with JBIC and one with Government of Australia); and (d) Monitoring and Evaluation Manual. Table 6 lists the major outputs of VAMESP II. The two joint evaluations with Japan in 2005 and 2007 were also deemed as pilot evaluations under VAMESP II.

Other donors have also supported Vietnam's ECD activities. Besides the two joint evaluations with Japan in 2005 and 2007 and the various activities of VAMESP (AusAID), the following programs/projects for capacity development in ODA management (including evaluation) have been implemented: (i) the Comprehensive Capacity Building Program for ODA Management (CCBP)²⁵ (Multi-donor assistance), (ii) Capacity Development of ODA Project Planning (CDOPP)²⁶ (JICA), (iii) Technical assistance on Enhancing ODA Absorptive Capacity and Efficiency²⁷ (ADB).

Therefore, thanks to donor supported M&E projects, the development of

²⁵ Various trainings with the aim of enhancement of ODA management capacity. Particular attention is paid to the implementation stage (monitoring).

²⁶ Training in project planning and IT with the aim of enhancing ODA management capacity of MPI and selected local government agencies.

²⁷ Training in project approval and implementation supervision and pilot monitoring with the aim of improving ODA disbursement.

M&E system in Vietnam was more advanced by the time of the 2007 evaluation as compared to the 2005 evaluation. This made the difference in performance and outcomes of the two joint evaluations in 2005 and 2007.

(3) Type of evaluation and methodology adopted

Thirdly, the 2007 MPI-JBIC joint evaluation used MPI's *Monitoring and Evaluation Manual: Evaluation Practice Module*, which was one of the outcomes of VAMESP II. The 2007 joint evaluation turned out to be the first full utilization of the *M&E Manual*.

The characteristics of the evaluation methods introduced in the *Evaluation Practice Module* can be summarized as follows:

- Based on principles such as cost effectiveness of evaluation, use of evaluation results in program/project management, participation of stakeholders, harmonization of evaluation methods, professionalized evaluation design, etc.
- Assess the value of the program/project from Plan-Actual Comparison and the DAC Five Evaluation Criteria (Relevance, Efficiency, Effectiveness, Impact and Sustainability).
- Use of the Logical Framework Approach for Plan-Actual Comparison.
- Preparation of the Evaluation Framework as a coherent tool throughout the evaluation process (design, data collection, data collation and analysis, conclusion).
- Use of various data collection methods such as literature review, questionnaires, interviews, direct observation, focus groups, etc. depending on evaluation questions. Use of both quantitative and qualitative data/information.
- Encouragement of focus groups and semi-structured interviews, which could enable evaluators to collect high-quality data for program/project impacts in relatively short time.
- Emphasis on analysis of factors behind success/failure of the program/project for drawing lessons and recommendations.

As already mentioned, the evaluation method in the *Manual* was developed with reference to the evaluation methods provided in the PCM handbooks of Japan. Aligning JBIC's ex-post evaluation method to Vietnam's was, therefore, relatively easy since there are many similarities between the evaluation frameworks of both Vietnam and Japan. In this sense, the relevance of MPI-JBIC joint ex-post project evaluation was high.

Regarding the 2005 MPI-MOFA joint evaluation, the first version (March

2003), of the *ODA Evaluation Guideline* of MOFA was used. But since the evaluation method for program-level evaluation was very new in Vietnam (even in Japan program-level evaluation is still under discussions), the theory as well as practice of program-based evaluation was difficult for the Vietnamese core-team members to grasp. Particularly for participants from the executing agencies and PMUs who are handling the respective projects in their day-to-day works, the methodology of project evaluation is easier to understand than that of program-level evaluation. If the 2005 MPI-MOFA joint evaluation employed a project-level evaluation approach, the impact on ECD might have been much more effective.²⁸

(4) Inputs from Japanese side

Fourthly, the inputs from Japan in terms of personnel and money were different between the two joint evaluations. On the one hand, the total budget for the 2005 MPI-MOFA joint evaluation was approximately 18 million Yen and the work volume of consultants totaled 10 Man-Months (M/M) (5 M/M for the Japanese evaluation experts and 5 M/M for the national consultants). On the other hand, the total budget for the 2007 MPI-JBIC joint evaluation was approximately 40 million Yen and the work volume of consultants totaled 17 Man-Months (M/M) (8 M/M for the Japanese evaluation experts and 9 M/M for the national consultants).

It is difficult to simply compare the size of budget and volume of personnel inputs between the two joint evaluations because the conditions for the type of evaluation, number of target projects, scope of works, etc. were different. But it is evident that the Japanese inputs in the 2007 MPI-JBIC joint evaluation were larger than in the 2005 MPI-MOFA joint evaluation.

Since both of the two joint evaluation were designed based on the learning by doing with the practical approach through the workshop, training, and collaborative activities, the length and period of OJT by Japanese evaluation experts influence the outcomes of the learning. In this sense, the Vietnam core-team members of the 2007 MPI-JBIC joint evaluation benefited more than the 2005 members.

(5) Continuity of the person in charge

Fifthly, the 2007 MPI-JBIC joint evaluation benefited by the past experience

²⁸ Needless to say, there is a demarcation of responsibility in ODA evaluation in Japan as MOFA is in charge of policy and program-level evaluation and the executing agencies such as JICA are in charge of project-level evaluation in general.

of the persons in charge who also joined the 2005 MPI-MOFA joint evaluation. In fact, the same team leaders from both the Vietnamese and Japanese teams engaged in both joint evaluations in 2005 and 2007. The Vietnamese core-team leader was the Deputy Director of Foreign Economic Department of MPI and the Japanese team leader was a Japanese evaluation consultant.²⁹ Also one officer of MPI and one officer of MOT as well as two national consultants of Japanese team continuously committed to the two joint evaluations.

Both team leaders tried to improve the effectiveness of ECD through the framework of the 2007 MPI-JBIC joint evaluation based upon the experiences and lessons learned from the previous 2005 MPI-MOFA joint evaluation. Not only the Vietnamese team leader but also the Japanese team leader had accumulated experience, knowledge, know-how and a network for implementing joint evaluation in Vietnam. In this sense, the Vietnam core-team members of the 2007 MPI-JBIC joint evaluation benefited more than the 2005 members.

3-3 Potentials for ECD in Vietnam

Based upon the experience of the two joint evaluations between GOV and Japan, the following fields can be identified as potentials for ECD in Vietnam.

(1) Capacity Development on Evaluation Management

Although Decree 131 does not clearly mentioned that all types of evaluation shall be outsourced to external human resources (such as consultants), it is implied. The outsourcers are the Project Owner (in case of Initial Evaluation), PMU (Mid-term and Terminal Evaluations) or Line Agency (Ex-post Impact Evaluation). The 2006-2010 M&E Framework plans to set up a mechanism where PMUs hire external evaluators (consultants), but that mechanism has not yet been materialized, and PMUs have to perform all M&E tasks for their projects.

So far the outsourcing of evaluation has not been practiced yet, but in the near future the Line agencies, Project owners, and PMUs should be responsible for the evaluation management by outsourcing the respective evaluation works through the process of preparation of evaluation plan, terms of reference, tender and contracting to the consultants, monitoring of their activities and quality control of the outputs. This field must be developed.

²⁹ The author of this chapter.

(2) Establishment of Cost Norms for Evaluation

Also when conducting the evaluation works by respective Line agencies, Project owners, and PMUs either through outsourcing or by own staff, the budget for evaluation is necessary. But at present it seems difficult for them to pursue the responsibility due to personnel and budget shortage. According to MPI, evaluation outsourcing is still difficult mainly because the Cost Norms have not been established yet.³⁰ The Action Plan for the 2006-2010 ODA-M&E Framework plans to establish the Cost Norms through MOF in 2008. Therefore, the establishment of Cost Norms for evaluation is urgently required.

4. Conclusion

To sum up what we have presented here, we will raise our general ideas about how to better promote assistance in ECD. As mentioned repeatedly, the key lies in the generating the demand in partner countries for it.

4-1 General Findings

We have discussed factors for the success of JBIC's ECD model and in particular the joint evaluation cases in Vietnam in 2-3 and 3-2, respectively. The factors for success are summarized into four points as below: institutional capacity development, involving the right organizations, alignment of evaluation methods/ procedures, and coordination.

(1) Institutional capacity development

ECD is effective when an institutional framework for evaluation is established. If no such framework exists, discussions/ workshops on country evaluation systems as part of evaluation seminars could raise awareness of institutionalization of evaluation and create demand for it.

(2) Involving the right organizations

At earlier stages of ECD, ODA planning/ coordinating agencies tend to have better understanding of- and higher interests in- evaluation than executing agencies. However, once involved, executing agencies participate actively in evaluation processes (sometimes even more actively than planning agencies).

30 There exist general guidelines for procurement of ODA consultants, but not for ODA evaluation consultants.

(3) Alignment of evaluation methods/ procedures

Transfer of international evaluation standards is always necessary. However, the degree of involvement of the partner country in joint evaluation is higher when a country evaluation system and methodology is applied. In such cases, the Japanese side needs to align its evaluation procedures to the partner country evaluation system.

One of the differences between the 2005 and 2007 joint evaluations in Vietnam was that the former used Japanese evaluation methods/ procedures while the latter followed those of Vietnam. As such, the Japanese side needed to align its standard evaluation procedures and reporting format with that of Vietnam. For the Japanese side, this adjustment was a new task because the Japanese procedures had applied to joint evaluations in the past and thus it had always been the partner country which had to follow the Japanese procedures. In the current direction in which ECD is meant to build the country's own evaluation system, as is the case of the 2007 joint evaluation in Vietnam, alignment responsibility lies with the Japanese side.

(4) Coordination

As joint evaluation involves various organizations on the side of both the Japanese and the partner country, coordination among those organizations is very important. Usually, ODA planning/ coordinating agencies are supposed to play the coordinating role. However, there are often manpower or organizational constraints in those agencies. From a viewpoint of institutional enhancement, the coordination capacity of ODA planning/ coordinating agencies must be improved. At the same time, local consultants could play an important role in coordination.

Among the three joint evaluation cases presented in this chapter, coordination was: poor in the Indonesia case; not very crucial in the 2005 MPI-MOFA joint evaluation in Vietnam (because the Japanese side took initiative); key to the success of the 2007 MPI-JBIC joint evaluation. With a vertical administrative structure such as the one in Vietnam, MPI's leadership in coordinating various participating organizations was indispensable. Also in this case, Vietnamese consultants hired by the Japanese evaluator played an important role in reminding MPI of the procedures and follow-up correspondences from MPI to participating organizations.

4-2 Recommendations

Based on the above, we will try to list up some hints to be kept in mind when

planning and practicing ECD assistance.

For Evaluation Seminars:

(1) Create and cultivate needs/ demands for international standard ECD by transferring knowledge of and skills in evaluation.

Training in internationally-applied evaluation methods could be the starting point of ECD. The hands-on training style could help them gain interest in the subject matter, and provide a clear image about what will be useful for them and what will not. Some preparatory work prior to seminar sessions could enhance participants' willingness to attend the seminar.

(2) Identify the current situation of country evaluation systems.

Discussions on participating countries' evaluation systems could identify areas where ECD assistance is needed.

(3) Let participants learn from precedent cases of other countries: have them acquire new ideas to improve their evaluation systems.

Exchange of ideas with predecessors could help participants visualize what is involved in putting ECD into practice. Partner countries explanations of their experiences are often more persuasive than lectures by Japanese speakers.

For Joint Evaluations:

(4) Conduct joint evaluation as a pilot for establishment of a country evaluation system.

As mentioned in Section 3, a realistic and effective institutional framework on evaluation could be drawn from trials and errors through pilot evaluations.

(5) Accumulate evaluation practices on the side of partner countries.

Once established, the partner country must operate the institutional framework. In initial phases of operation of the institutional framework, however, involvement of donor-side (i.e., international) evaluators through joint evaluation could be helpful to accumulate good evaluation practices.

(6) Involve executing agencies to raise their awareness of evaluation.

There is a tendency that although executing agencies are in a position to conduct project evaluation, they tend to concentrate on implementation of ongoing projects instead of the evaluation of such ongoing or completed pro-

jects. Executing agencies could perceive the significance of evaluation if they join the joint evaluation team and learn that the evaluation findings could improve their project management.

(7) Start from joint evaluation of projects (i) for which the partner government has spent a large amount of their own resources on or (ii) whose impact was large.

Ownership of evaluation is high if ownership of the evaluated project is high. This is why we consider that joint evaluation of ODA loan projects interest partner countries more than grant aid, and thus is most likely a good entry point for evaluation.

(8) Conduct joint evaluation for alignment and harmonization of evaluation methods and procedures to the partner country evaluation system.

Alignment and harmonization of aid procedures should cover evaluation. Joint evaluation would be a good opportunity for both donor and partner countries to pursue alignment in the area of evaluation.

For Institutional Enhancement:

(9) Work on institutionalization of evaluation after the demands for country evaluation systems are somewhat understood.

As mentioned above, evaluation seminars/ training and joint evaluation practices can raise demand for ECD, and make it easier to promote institutionalization of evaluation.

(10) Work with the right and capable organizations.

As evaluation is still a new concept in many countries, in addition to working on increasing demand, institutionalization needs a strong commitment of leaders to promote evaluation.

(11) Widen the resource base for evaluation from the government sector to include the private sector and academics.

When building a country evaluation system, it is better to consider developing an institutional framework of outsourcing of evaluation as well. This is particularly true when government human resources cannot be allocated to evaluation works such as data collection, analysis and report-writing, or where external (third-party) evaluation is preferred. The framework might

include the development of cost norms and training of non-governmental evaluation human resources.

Annex Table 1: details of evaluation seminars

Annex Table 1-1. Details of Evaluation Training ODA Loan Projects (Example of 2007 Seminar)

Time	Program
October 18 (Thu) 10:00-12:30	♦ Module 1: Introduction and Overview of JBIC ODA Loan Project Evaluation <ul style="list-style-type: none"> • Introduction of Moderators • Sharing of participants' experience in project monitoring and evaluation • Reminder of Procedure of JBIC ODA Loan Project Evaluation • Introduction to the case study "Karnac Tunnel Project" with a small exercise on Project Outline
October 22 (Mon) 10:00 - 12:30	♦ Module 2: Evaluation of Relevance <ul style="list-style-type: none"> • Lecture • Group discussion on Worksheet 1 of Pre-departure Exercise
	♦ Module 3: Evaluation of Efficiency <ul style="list-style-type: none"> • Lecture • Group discussion on Worksheet 1 of Pre-departure Exercise
14:00 - 16:30	♦ Module 4: Evaluation of Effectiveness <ul style="list-style-type: none"> • Lecture • Group discussion on Worksheet 1 of Pre-departure Exercise • Group discussion on Worksheet 2 of Pre-departure Exercise
October 23 (Tue) 10:00 - 12:30	♦ Module 5: Evaluation of Impact <ul style="list-style-type: none"> • Lecture • Group discussion on Worksheet 1 of Pre-departure Exercise • Review of Sample Beneficiary Survey Report
14:00 - 16:30	♦ Module 6: Evaluation of Sustainability <ul style="list-style-type: none"> • Lecture • Group discussion on Worksheet 1 of Pre-departure Exercise
October 24 (Wed) 10:00 - 12:30	♦ Module 7: Evaluation Feedback (Lessons Learned, Recommendations and Rating) <ul style="list-style-type: none"> • Lecture • Group discussion on Worksheet 1 of Pre-departure Exercise • Review of Sample Ex-post Evaluation Report • Wrap-up

Source: Materials for ODA Project Evaluation Seminar, JBIC, 2007.

Annex Table 1-2. Details of Evaluation Systems Workshop Modules (Example of 2007 Seminar)

Time	Program
October 24 (Wed) 14:00 - 16:30	♦ Guidance on Evaluation Training Plan and Final Report <ul style="list-style-type: none"> • Purpose and reporting format of Training Plan and Final Report • Possible ways to utilize what participants will have learned from this Seminar
	♦ Discussion on effective ways of disseminating seminar contents in participating countries <ul style="list-style-type: none"> • Seminar contents worth disseminating and those that are not • Possible constraints in disseminating seminar contents • Other topics (as needed)
	♦ Individual work to prepare and submit the Training Plan

- October 25 (Thu) ♦ Joint Evaluation Case Study**
- 09:30 - 12:30**
- Guest speakers' presentations on their joint evaluation experiences
 - Feedback from seminar participants
- 14:00 - 16:30 ♦ Joint Evaluation Case Study (continued)**
- ♦ Discussion on Joint Evaluation as a way of evaluation capacity development**
- Advantages and disadvantages
 - Opportunities and constraints
 - Respective roles of evaluation seminars and joint evaluation in the common context of evaluation capacity development
 - Wrap-up discussion, recommendations for JBIC
- 16:30-17:30 ♦ 2nd discussion on the Training Plan**
- ♦ Brush up the individual Training Plan (Deadline of the submission is 26th)**

Source: Materials for ODA Project Evaluation Seminar, JBIC, 2007.

Annex Table 1-3. Problems and Measures on Evaluation Systems in Countries Participating in ODA Loan Project Evaluation Seminars (2004, 2005 and 2006)

Evaluation stage	Problems	Measures against problems
General	<ul style="list-style-type: none"> • Lack of ODA evaluation system and legal basis for it • Lack of evaluation budget • Inability to collect information for evaluation (e.g. inadequate statistics, insufficient capability of evaluators to collect information) • Scarcity of local human resources in evaluation • Preparation of evaluation reports only by donor side 	<ul style="list-style-type: none"> • Introduction of legislation for ODA evaluation • Harmonization/ alignment of evaluation methods • Preparation of evaluation guidelines by developing countries • Evaluation trainings/ seminars funded by donors (training of local trainers) • Joint evaluation practices
Ex-ante evaluation	<ul style="list-style-type: none"> • Poor conduct of ex-ante evaluation • Focus only on feasibility, not on assessment of outcomes/ impacts • No or unrealistic indicators/ targets • Evaluation only on engineering aspects • Non-involvement of third party in evaluation • Insufficient feedback of past experience to planning of new projects 	<ul style="list-style-type: none"> • Strengthening of feasibility studies, thereby setting of realistic indicators/ targets • Formation of multidisciplinary evaluation team
Monitoring/ mid-term evaluation	<ul style="list-style-type: none"> • Lack of communication among concerned agencies in developing countries • Unclear role of governments of developing countries in monitoring and evaluation (in relation to on-going contracts with contractors and consultants) • Lack of data on project progress • Monitoring of disbursement only 	<ul style="list-style-type: none"> • Sharing of incentives for monitoring among concerned agencies in developing countries • Strengthening of ex-ante evaluation (for good implementation) • Development of project management information system

Ex-post evaluation	<ul style="list-style-type: none"> • Insufficient conduct of ex-post evaluation as part of country evaluation system (only following evaluation by donors) • No mention of conduct of ex-post evaluation in project agreements • Lack of recognition of executing agencies by necessity and procedure of ex-post evaluation 	<ul style="list-style-type: none"> • Conducting of ex-post evaluation by developing countries themselves with support from donors • Clear mention of conduct of ex-post evaluation in project agreements • Selection of evaluators by developing countries
--------------------	--	---

Source: Materials for ODA Project Evaluation Seminar, JBIC, 2007.

Annex Table 2: Details of Joint Evaluation Programs in Vietnam

Annex Table 2-1. Japanese ODA Projects under the Red River Transport Development Program (1994-2004)

No.	Sub-Sector	Type of Aid	Project	Donor	Year
01	road	loan	National Highway No.5 Improvement Project (1)(2)(3)	JBIC	1996-2004
02	road	loan	[PHASE I] National Highway No.1 Bridge Rehabilitation Project (1)(2)(3)	JBIC	1996-2005
03	road	loan	[PHASE II] National Highway No.1 Bridge Rehabilitation Project (1)(2)(3)	JBIC	1999-2004
04	road	loan	National Highway No.10 Improvement Project (1) (2)	JBIC	1998-2007
05	road	loan	National Highway No.18 Improvement Project (1) (2)	JBIC	1998-2008
06	road	loan	Bai Chay Bridge Construction Project	JBIC	2001-2008
07	road	loan	Binh Bridge Construction Project	JBIC	2000-2007
08	road	loan	Red River (Thanh Tri) Bridge Construction Project (1)(2)(3)	JBIC	2000-2008
09	road	loan	Transport Infrastructure Development Project in Hanoi	JBIC	1999-2006
10	road	grant aid	Project for Reconstruction of Bridges in the Northern District	MOFA/ JICA	1996-1998
11	road	grant aid	Project for Improvement of Transport Technical and Professional School No.1 in Vietnam	MOFA/ JICA	2000
12	road	technical coop. project	Project for Strengthening Training Capabilities for Road Construction Workers in Transport Technical and Professional School No.1 in Vietnam	JICA	2001-2006
13	road	development studies	Feasibility Study of the Highway No.18 Improvement in Vietnam	JICA	1995-1996
14	road	development studies	Study on Urban Transportation for Hanoi City in Vietnam	JICA	1995-1996
15	road	development studies	Detailed Design of the Red River Bridge (Thanh Tri Bridge) Construction Project	JICA	1998-2000

16	road	development studies	Vietnam National Transport Development Strategy Study (VITRANSS)	JICA	1998-2000
17	railway	loan	Hanoi-Ho Chi Minh City Railway Bridge Rehabilitation Project (1)(2)(3)	JBIC	1994-2005
18	railway	development studies	Upgrading the Hanoi-Ho Chi Minh Railway Line to Speed up Passenger Express Trains to an Average Speed of 70 km/h	JICA	1993-1995
19	port & sea	loan	Hai Phong Port Rehabilitation Project (1) (2)	JBIC	1994-2007
20	port & sea	loan	Cai Lan Port Expansion Project	JBIC	1996-2005
21	port & sea	loan	Costal Communication System Project	JBIC	1997-2002
22	port & sea	technical coop. project	Project on Improvement of Higher Maritime Education in Vietnam	JICA	2001-2004
23	port & sea	development studies	Feasibility Study for Construction of Cai Lan Port	JICA	1993-1994
24	port & sea	development studies	Master Plan Study of Coastal Shipping Rehabilitation and Development Project	JICA	1994-1996
25	inland waterway	development studies	Study of Red River Inland Waterway Transport System in Vietnam	JICA	2001-2003

Source: Final Report of Vietnam-Japan joint Evaluation on the Japanese ODA Program for the Transport Infrastructure Development in the Red River Delta Area of the Socialist Republic Vietnam, MPI & MOFA (February 2006)

Note:

- 1) A series of ODA loan projects with separate loan agreements such as the phased project for NH 1, 5, 18, Red River bridge, Hai Phong port, etc. are deemed as one project in the study for convenience.
- 2) The actual project area of the "Hanoi-Ho Chi Minh City Railway Bridge Rehabilitation Project (1)(2)(3)" is the central part of Vietnam, however considering the linkage between the project and the Red River Delta in terms of the outcome of the Program, this project is included as one of the components of the Program.
- 3) Through the initiative of the JICA Vietnam Office, a relatively small "Traffic Safety Promotion Program I (2002) & II (2003-2004)" and the "Basic survey on road traffic safety in Hanoi city (2003-4)" were executed.
- 4) "The Master Plan Study on the Transport Development in the Northern Part in the Socialist Republic of Vietnam" by JICA 1993-1994 was the original plan for the subject of this joint evaluation survey, so this is not included in this list.

Annex Table 2-2: Comparison of Two Joint Evaluations in 2005 and 2007

	MPI-MOFA Joint Evaluation in 2005	MPI-JBIC Joint Evaluation in 2007
Type of Evaluation	Program-level evaluation	Ex-post project evaluation
Target Projects	13 ODA loan projects, 2 grant aid projects, 2 technical cooperation projects, and 8 development studies in transport sector	3 ODA loan projects in transport sector
Implementation Schedule	July 2005–February 2006	August 2007–June 2008
Evaluator	<p><Vietnam core-team></p> <ol style="list-style-type: none"> 1. Ministry of Planning and Investment (MPI) • Foreign Economic Dept. (3), Infrastructure Dept. (1) 2. Ministry of Transport (MOT) • Planning and Investment Dept. (2), PMU 18 (1), TDSI (1) 3. VAMESP II (2) <p><MOFA team></p> <ol style="list-style-type: none"> 1. Ministry of Foreign Affairs (MOFA) • Embassy (2), Evaluation Division (2) 2. Consultants • Japanese evaluation experts (2), National consultants (3) 	<p><Vietnam core-team></p> <ol style="list-style-type: none"> 1. Ministry of Planning and Investment (MPI) • Foreign Economic Dept. (2), Appraisal Dept. (2), Infrastructure and urban development Dept. (2) 2. Ministry of Transport (MOT) • Planning and Investment Dept. (2), PMU 5 (2), PMU 18 (2), Railway PMU (3), Road Administration Authority (1), Railway Administration Authority (1), Vietnam Railway Corporation (1), PMU on Transport Safety (NTSC) (1), TDSI (2) <p><JBIC team></p> <ol style="list-style-type: none"> 1. Consultants • Japanese external evaluator (2), National consultants (2)
Evaluation Method and Tools	<p>ODA Evaluation Guideline (First Version) (March 2003, MOFA)</p> <ul style="list-style-type: none"> • Consistent with OECD-DAC guideline • Combination of multiple survey method including the questionnaire survey with supplementary interview, interview survey, beneficiary survey with Semi-Structured interview, direct observation by project site visit, and document review 	<p>Monitoring and Evaluation Manual: Evaluation Practice Module (May 2007, MPI)</p> <ul style="list-style-type: none"> • Based on five evaluation criteria of OECD-DAC • Combination of multiple survey method including the questionnaire survey with supplementary interview, interview survey, beneficiary survey with Semi-Structured interview and Focus Group, direct observation by project site visit, and document review • Use of economic evaluation (EIRR) • Introduction of evaluation rating
Inputs from Japan	<p>Japanese consultants: 5 M/M</p> <p>National consultants: 5 M/M</p> <p>Total budget: Approx. 18 million JPY</p>	<p>Japanese consultants: 8 M/M</p> <p>National consultants: 9 M/M</p> <p>Total budget: Approx. 40 million JPY</p>

Division of Role/Demarcation of Tasks at Each Stage	MPI-MOFA Joint Evaluation in 2005		MPI-JBIC Joint Evaluation in 2007	
	MOFA Team	Vietnam Core-team	JBIC Team	Vietnam Core-team
Stage 1: Evaluation Planning				
(1) Evaluation workshop and training	<ul style="list-style-type: none"> Plan and implement the workshop (2 days) 	<ul style="list-style-type: none"> Participate in the workshop (2 days) 	<ul style="list-style-type: none"> Plan and organize the workshop and post-workshop training (6 days) 	<ul style="list-style-type: none"> Participate in the workshop and post-workshop training (6 days)
(2) Evaluation framework	<ul style="list-style-type: none"> Prepare the draft objective framework and the draft evaluation framework Finalize the objective framework and evaluation framework based on the comments from VN team 	<ul style="list-style-type: none"> Review the draft objective framework and draft evaluation framework prepared by MOFA team and provide comments 	<ul style="list-style-type: none"> Prepare the draft logframe and draft evaluation framework Jointly finalize the logframe and evaluation framework jointly 	<ul style="list-style-type: none"> Review and jointly finalize the logframe and evaluation framework
(3) Questionnaire	<ul style="list-style-type: none"> Prepare the questionnaires 	<ul style="list-style-type: none"> Send the finalized questionnaires to each informant 	<ul style="list-style-type: none"> Prepare the draft questionnaires except to the project executing agencies Finalize all of the questionnaires 	<ul style="list-style-type: none"> Prepare the draft questionnaires to the project executing agencies Send the finalized questionnaires to each informant
(4) Pilot beneficiary survey			<ul style="list-style-type: none"> Plan and demonstrate the pilot beneficiary survey 	<ul style="list-style-type: none"> Participate the pilot beneficiary survey
(5) Field survey plan	<ul style="list-style-type: none"> Plan the draft field survey plan Finalize the filed survey plan 	<ul style="list-style-type: none"> Review and comment on the filed survey plan 	<ul style="list-style-type: none"> Jointly prepare and finalize the draft filed survey plan 	<ul style="list-style-type: none"> Jointly prepare and finalize the draft filed survey plan
Stage 2: Data Collection and Analysis				
(1) Field survey and data collection	<ul style="list-style-type: none"> Jointly arrange interview appointments Jointly implement the interview survey to the informant, beneficiary survey and project site visits Jointly collect the answered questionnaires Conduct economic impact study 	<ul style="list-style-type: none"> Jointly arrange interview appointments Jointly implement the interview survey to the informant and project site visits Jointly collect the answered questionnaires 	<ul style="list-style-type: none"> Jointly arrange appointments Jointly implement the interview survey to the informant, beneficiary survey (SSI and Focus Group) and project site visits Jointly collect the answered questionnaires 	<ul style="list-style-type: none"> Jointly arrange appointments Jointly implement the interview survey to the informant, beneficiary survey (SSI and Focus Group) and project site visits Jointly collect the answered questionnaires

Division of Role/Demarcation of Tasks at Each Stage	MPI-MOFA Joint Evaluation in 2005		MPI-JBIC Joint Evaluation in 2007	
	MOFA Team	Vietnam Core-team	JBIC Team	Vietnam Core-team
(2) Data analysis	<ul style="list-style-type: none"> • Compile and analyze the collected data and information 	<ul style="list-style-type: none"> • Share the compiled data and information by MOFA team 	<ul style="list-style-type: none"> • Compile the collected data and information • Analyze the collected data and information 	<ul style="list-style-type: none"> • Partially compiled data and information • Partially analyze the collected data and information regarding the efficiency and effectiveness criteria
Stage 3: Conclusion of Evaluation				
(1) Draft Evaluation Summary Results	<ul style="list-style-type: none"> • Prepare the preliminary evaluation results including the lessons learned and recommendations • Revise the preliminary evaluation results based on the comments of Vietnam Core-team 	<ul style="list-style-type: none"> • Review the preliminary evaluation results by MOFA team and provide comments 	<ul style="list-style-type: none"> • Prepare the 1st draft evaluation summary results (impact, sustainability) • Review and comments on the draft evaluation summary results by Vietnam Core-team • Review and comments on the draft the lessons learned and recommendations • Review and comments on the draft evaluation summary results by JBIC team • Jointly prepare the 2nd draft evaluation summary results 	<ul style="list-style-type: none"> • Prepare the 1st draft evaluation summary results (relevance, efficiency and effectiveness) • Draft the lessons learned and recommendations • Review and comments on the draft evaluation summary results by JBIC team • Jointly prepare the 2nd draft evaluation summary results
(2) Internal Feedback			<ul style="list-style-type: none"> • Jointly finalize the evaluation summary results based on the comments from the participants of the internal feedback seminar 	<ul style="list-style-type: none"> • Presentation the 2nd evaluation summary results at the internal feedback seminar • Jointly finalize the evaluation summary results based on the comments
(3) Evaluation Report (Final Report)	<ul style="list-style-type: none"> • Prepare the draft evaluation report • Finalize the evaluation report 	<ul style="list-style-type: none"> • Review and comments on the draft evaluation report 	<ul style="list-style-type: none"> • Prepare the draft evaluation report • Finalize the evaluation report 	<ul style="list-style-type: none"> • Review and comments on the draft evaluation report

Division of Role/Demarcation of Tasks at Each Stage	MPI-MOFA Joint Evaluation in 2005		MPI-JBIC Joint Evaluation in 2007	
	MOFA Team	Vietnam Core-team	JBIC Team	Vietnam Core-team
Stage 4: Feedback				
(1) Feedback seminar			<ul style="list-style-type: none"> Plan and organize the feedback workshop Present the achievement of joint evaluation at the feedback workshop Release the evaluation report on JBIC's web site 	<ul style="list-style-type: none"> Present the evaluation results at the feedback workshop Press release to the media
(2) Publication of Evaluation Results	<ul style="list-style-type: none"> Release the evaluation report on MOFA's web site 	<ul style="list-style-type: none"> Publish in Monitoring and Evaluation Manual: Evaluation Practice Module (May 2007, MPI) as a case study of program evaluation 		<ul style="list-style-type: none"> Release the evaluation report on MPI's web site

Source: Prepared by the author.

4

Evaluation Capacity Development in the Asia-Pacific Region: A Proposal for an Asia-Pacific Evaluation Association Network (APEA NET)

Ryokichi Hirono

1. Introduction

Development evaluation is increasingly recognized today as critical for more efficient, effective and accountable socio-economic development, resulting in a greater emphasis on better evaluation practices and institutional strengthening at corporate, local and national levels in all countries of Asia and the Pacific region. They are now being steadily woven into the social governance system in an increasing number of countries in the region. Simultaneously, many studies on evaluation theories, methodologies, approaches, and systems have been conducted by scholars and researchers, and several countries in this region have already established country-based Evaluation Societies.

Evaluation, however, is a comparatively new profession in this region where evaluation culture does not have sufficient roots among individuals and organizations, particularly in the public sector. For this reason, it seems essential that central and local governments and other development stakeholders accelerate their efforts for strengthening evaluation culture among the public, developing evaluation experts, mainstreaming development evaluation in their national and local socio-economic development plans, promoting the exchange of evaluation information and experiences among evaluation professionals, networking among the existing national evaluation societies, and, where no evaluation societies exist, encouraging the establishment of national evaluation societies. Such efforts will undoubtedly contribute to the further advancement of evaluation theories and practices and to

more efficient and effective socio-economic development in Asian and Pacific countries. It was against this background that development-oriented evaluation experts organized the International Development Evaluation Association (IDEAS) ¹ in Beijing in 2002 and parties generally concerned with evaluation organized the International Organization for Cooperation in Evaluation (IOCE) ² in Lima in 2003. Also against this background, participants at the Asia-Pacific meeting on the Paris Declaration held at the Asian Development Bank in Manila three years ago, jointly proposed the establishment of the ASIA-PACIFIC EVALUATION ASSOCIATION NETWORK (APEA NET). All country delegations participating in the meeting including the Japan Evaluation Society (JES) ³, the Sri Lanka Evaluation Association (SLEvA) ⁴, the Malaysian Evaluation Society ⁵ and participants interested in promoting evaluation culture in the region welcomed the proposal. ⁶

In this paper the author, drawing on observations of evaluation systems and practices in the Asia-Pacific region during the last few decades, will delineate various challenges facing developing countries in the region in regards to development and development evaluation. Also, giving due recognition to the urgent need in these countries and the repeated call in bilateral and multilateral donors for improving development evaluation to enhance aid and development effectiveness, the author presents some thoughts based on preliminary discussions carried out with evaluation experts in and outside the Asia-Pacific region, on the objectives, possible programs, membership, management and financing of the APEA NET.

2. A Growing Interest in Establishing National Evaluation Societies in Developed Countries

Since the 1980s, country after country in the West began to establish national evaluation societies including the American Evaluation Association, the Canadian Evaluation Association, the French Evaluation Society, the German Evaluation Society, the Italian Evaluation Society, and the United Kingdom Evaluation Society. ⁷ According to a 2003 survey commissioned by the Ministry of Foreign Affairs, the Government of Japan, most of the governments (and the national evaluation societies and international organizations) surveyed have been shifting their priority evaluation exercises from project to program/sector, policy and country evaluation. In recent years they have all been engaged in evaluating and reviewing national government policies related to Millennium Development Goals (MDGs) set by the United Nations

General Assembly in the fall of 2000.⁸ Japan Evaluation Society established in 2000 was the last among G-7 countries that organized a national evaluation society. Bilateral and multilateral donors, interested in promoting evaluation culture and enhancing aid and development effectiveness in partner countries, have been facilitating the mushrooming in the number of national evaluation societies in developing countries since the 2000s. Particularly in developed countries during the last two decades or so, growing interest in establishing national evaluation societies was precipitated by several common factors.

First, throughout the 20th century the separation of corporate management from corporate ownership (through capital market development and pressures from economic globalization involving domestic deregulation and foreign trade, investment, and finance liberalization) had contributed tremendously to the rapid expansion of the corporate world accompanied by management modernization. Corporate ownership also began to undergo enormous changes in the latter half of the 20th century including a shift from individual to institutional ownership as represented by pension and investment funds. All this resulted in an environment in which shareholders increasingly demand corporate management to be transparent and accountable to their owners. The demand coincided with an increasing concern shown by governments of Western countries to standardize corporate accounting and auditing procedures and practices (as corporations became more multinational both in terms of ownership and business operations) to minimize tax evasion through transfer pricing practices from high to low taxing countries. All these socio-economic changes gave rise to: an increased need for corporate accountants and auditors; the phenomenal expansion of professional schools and institutes to meet the increased demand for qualified personnel; and the establishment of their national associations to safeguard their professional interests and societal standing. Evaluation culture in Western societies thus began to emerge in the increasingly multinational corporate world in terms of improved and standardized accounting and auditing practices, triggered by the changing nature of corporate activities and management practices as well as to the government regulatory policies including taxation.

Second, in the latter part of the 20th century government activities became more extensive, corresponding to corporate activities which had become: increasingly multinational: far reaching in terms of their impact on communities and national economies: ever more complex. In response to changing socio-economic environments at home and overseas, government

activities, through dynamic formulation and implementation of macro-economic and -social security policies, started to go far beyond the traditional confines of national safety and security move into a whole range of economic and social security of the people. With the people's growing positive expectations toward government actions to safeguard their economic, social and political rights, people increasingly began to demand legislative, executive and judiciary actions to improve the transparency and accountability of government activities in running public administration as well as in policy formulation and implementation. In response to these popular pressures, governments of Western countries had to improve their own accounting and auditing procedures and practices in addition to installing monitoring and evaluation procedures and practices in all their public administration activities which had an institutionalizing effect in all departments/ministries and agencies and in the reporting of evaluation results to the public through parliamentary debates and the mass media. All this culminated in an increased demand for professional evaluators; a phenomenal expansion of professional schools and institutes to develop and train high-quality evaluators; and the establishment of national evaluation societies.

Third, although the international community had been increasing its official development assistance to developing countries consistently during the postwar period 1947-1990 mainly on ideological/political grounds, ODA growth was restrained during the 1990s in response to the end of the Cold War, budgetary constraints in major donor countries and a growing concern in regards to aid effectiveness as a whole. Since 2001, global ODA has only just begun to increase again mainly on humanitarian and anti-terrorism grounds. While welcoming the quantitative expansion of bilateral and multi-lateral ODA, the 1990s has witnessed a persistent and growing concern, both in donor and recipient countries over corruption in aid practices and aids' effectiveness in terms of its implementation and outcome. This concern culminated in 2004 in the OECD/Development Assistance Committee (DAC) adopting the Paris Declaration on Aid Effectiveness. The Paris Declaration, including its 2008 assessment in Ghana, emphasizes the necessity of the international community to enhance efforts in regards to Ownership, Alignment, Harmonization, Management for Results and Evaluation. Reinforcing the Paris Declaration, the Hanoi Statement in 2006 gave top priority to capacity building in developing countries for both policy (policy formulation, implementation, and monitoring) as well as evaluation. The need for improving aid and development evaluation and strengthening evaluation

capacity in developing countries has thus been recognized as being crucial for the success of development, in general, and, in particular for achieving the United Nations Millennium Development Goals (MDGs).

Thus, auditing and its expanded and reformed version, evaluation is a product of the tension (often confrontation) between shareholders (owners, investors), bondholders and financing institutions (lenders, investors), on the one hand, and company management at the corporate level on the other; and between the people and their representative parliament and assemblies, on the one hand, and the executive branch of the government at the local and central government level on the other. At both levels, however, it has been the corporate shareholders and lenders as well as the public and their representative assemblies that are demanding a greater degree of transparency and accountability for improved corporate management and public administration. Furthermore, with an increasing emphasis in recent years on corporate social responsibility, corporate evaluation has to respond not only to the demands from immediate interest groups such as shareholders and lenders but also to the demands from consumers and communities, as shown in cases related to food safety problems as well as illicit branding. In addition, with ongoing progress in economic globalization, the evaluation of government activities and public administration has to respond not only to the general public and parliaments/assemblies at the national and local levels, but also to the people and governments overseas, i.e., the international community. Today, therefore, all stakeholders, in both developed and developing countries, increasingly share concerns in regards to the monitoring and evaluation of corporate and government activities, including formulation and implementation of project, program, and policy.

3. Major Challenges of Development Evaluation in Developing Countries in General and Asia-Pacific Region in Particular

1) Major Issues of Development

As shown in national and regional (sub-national) socio-economic development plans, developing countries in recent decades have been more or less confronted with a number of major development issues. As shown in Table 1, while differences still exist among developing countries and regions in the achievements of the Millennium Development Goals (MDGs) made so far, the Millennium Declaration and the MDGs signify the major development

issues currently facing developing countries in general including the problems of HIV/AIDS and other infectious diseases. The most pressing major issues that have been confronting countries in all developing regions and will foreseeable future continue to confront them are: a) restraining population expansion and reducing dependency rate, b) increasing per capita gross national income/products through a steady and, if possible, accelerated economic growth, c) reducing poverty and the growing income gap, as well as high levels of unemployment and/or underemployment, d) narrowing the increasing income, social and opportunity gaps between different regions (sub-national) and segments of the population as well as gender inequality, e) environmental deterioration in air, water and soil (See Tables 2 & 3). Furthermore, most developing countries, including those in the middle income group, that have been struggling to sustain their economic growth, urgently need the following: f) increasing the relatively low level of capital formation as a percent of GDP, with a view to expanding infrastructure, particularly for power and drinking water supply, transportation and communication, g) expanding exports to reduce growing trade deficits, while reducing their excess dependence on official development assistance from the international community, and h) improving fiscal balance through higher efficiency of public administration and better governance including tightened anti-corruption measures. (See Table 4)

Every year since 2001 the United Nations⁹ and the World Bank¹⁰ have been publishing progress reports on the MDGs (their most recent report was the 2008 version). The European Commission also published a background paper for the forthcoming European Report on Development in September 2008 titled "Millennium Development Goals at Midpoint: where do we stand and where do we need to go? ".

The findings of these three reports from three different international organizations can be summarized as follows.

- i) Most countries of South Asia and Sub-Saharan Africa will fail to achieve most goals specified in the MDGs, while most countries of East Asia and the Pacific as well as Latin America and the Caribbean may achieve their MDGs (with some exceptions of the least developed countries in the region);
- ii) With a slowdown in the global economy precipitated by the sub-prime mortgage crisis in the U.S., together with oil and food price shocks, many developing countries hitherto on the right track of achieving the MDGs may also find achieving some of their MDGs more diffi-

cult, due to growing import restrictions, less direct investment, less aid and other financial flows, including worker remittance from developed countries;

- iii) Developing countries, particularly countries finding it difficult to achieve their MDGs may need to give a greater emphasis on policy coherence between macro-economic policy and sectoral policies. These countries will also have to ensure policy coherence between different sectoral policies so that it will result not only in sustained economic growth but also well-integrated and well-balanced sectoral developments, contributing inter alia to strengthened economic institutions and favorable business environments;
- iv) As both developing and developed countries are confronted with an acute downturn in growth prospects and rising unemployment and under-employment, prospects over the next few years for achieving MDGs are increasingly pessimistic. In order that the progress made so far in the achievement of the MDGs does not retrogress over the next few years, the international community needs to focus their assistance on strengthening basic human needs and social safety-nets for the poorest of the poor within and across countries.
- v) When assisting fragile states in the achievements of the MDGs, the special needs of fragile states will have to be taken into account more seriously; and
- vi) Improved global economic governance will help many developing countries to achieve their MDGs, including the progress of the Doha Rounds, better regulation of their financial systems, reduction in the barriers to unskilled labor migration and the mitigation of the risk of global warming, as well as the enhanced adaptation activities in developing countries.

It is important to recognize, however, that “Notwithstanding the issues raised by the MDGs, they have played an essential political role in mobilizing the support for development assistance at a time when aid disbursements were on a downward trend in many key OECD countries.”¹¹

Countries in Asia and the Pacific region, have generally been performing better than those in other regions of the developing world in terms of economic growth, employment, education and primary healthcare. Countries in Asia and the Pacific region, have not done as well on the issues of social and gender inequality, corruption, inefficiency of public administration and inadequate governance as well as environmental destruction. Not only have these

issues remained unresolved but in some countries, even under a high pace of economic growth, conditions have worsened over the last half a century. In Asia-Pacific countries, with the possible exception of Singapore, relatively rapid economic growth over a prolonged period has not solved these critical social issues. (See Tables 5 and 6)

2) Emergence of National Interest and Concern with Development Evaluation

It is against these backgrounds that countries in the region together with their bilateral and multilateral donors have shown a growing interest in development evaluation. It is fair to say, however, that pressure on both bilateral donors as well as stakeholders and not the initiatives of developing countries (Asia-Pacific partner countries) themselves that has caused this interest in development evaluation which contributed to the introduction in the 1990s of aid and development evaluation procedures and practices. Whereas bilateral donors are under pressures from their own tax payers and other stakeholders at home, international organizations such as the United Nations Development Programme (UNDP), the World Bank and the Asian Development Bank (AsDB) are under pressure from their respective major donors/contributors.

It was probably too early, at that time, for developing countries in the region to generate an evaluation culture due to the community social structures held from one generation to another for many centuries and their concomitant political regimes where people had been either accustomed not to ask questions or dependent upon and subservient to the heads of their own communities and/or to the State. Incidentally, Asian communalism is often said to be an easy entry to Asian communism. It could also be said that economic globalization with its emphasis on marketization and structural adjustment policies, i.e., deregulation of domestic economic activities and external trade and investment liberalization, and the consequent political and social globalization with its emphasis on smaller and more efficient central government machinery, decentralization and devolution of authority, transparent and accountable governance and rights-based approach to socio-economic development involving a greater participation of civil society have all precipitated developing countries in general and Asia-Pacific countries in particular to respond positively over time to the emergence of an evaluation culture and a growing interest in evaluation per se and especially development evaluation.

There is no doubt, however, that, together with the rise of civil society movement at home, the growing budgetary deficits in many of these countries has been an immediate factor responsible for the governments of these Asia-Pacific countries to better understand the need for installing development evaluation in the public sector (first at the project level, then to the program level and eventually to the policy level). Governments in the region today increasingly feel the vital necessity to scrutinize and better manage every public expenditure (including loans and grants from bilateral and multilateral donors). Such scrutiny and better management is meant to minimize the development cost in all sectors of their economies and every development project, program and/or policy and therefore maximize benefits accruing from such development interventions at local and national levels.

3) Challenges of Development Evaluation in the Region

Confronted by these urgent development tasks, the challenges of development evaluation have been enormous and diverse in many Asia-Pacific countries, simply due to the late development of the evaluation culture and more often than not, due to the under-development of evaluation experts and staff and the inadequate development of the information and data collection and analysis system required for quality evaluation. It is apparent that the belated development, in these countries, of evaluation policy and programs including evaluation guidelines and manuals resulted essentially from policy-makers' inadequate understanding of the critical importance of development evaluation in promoting efficient and effective development of communities, regions (sub-national) and the country. The less-than-desirable commitment, if not a lingering or apparent resistance, to evaluation among the highest political masters particularly at the program and policy level could be explained by their misconstrued interpretation and belief that any evaluation might be a threat to their established authority over development policy and program rather than an effort to assist them to improve their policy and program outcome and effectiveness.

It must be recognized, however, that some developing Asia-Pacific countries have gone ahead in setting up national evaluation machinery (China, India, Malaysia, Pakistan, the Philippines, Singapore, Sri Lanka, Thailand, and Vietnam). Bangladesh, India, Malaysia, Nepal, Pakistan and Sri Lanka have established national evaluation societies. Common factors appear to have been observed for such early successes. As discussed in the Section 2, these are first the increasing complexity of national development plan formu-

lation and implementation resulting from the need for achieving multiple development objectives of not only macro-economic growth and structural changes but also improved social wellbeing of the people, better distribution of growth benefits among different regions and segments of the population and environmentally sustainable development. Secondly, the increasing pressures of international competitiveness in all sectors of national economy under the on-going process of economic globalization have forced both governments and the private sectors to minimize the cost per unit of output, constantly improve the quality of products in terms of designs, safety and environmental sustainability and also meet fast and on-time delivery requirements.

Thirdly and related to the second, the size and the volume of national development expenditure financed by their own domestic financing have become much larger in the process of sustained economic and social development, precipitating their governments to look more closely into the efficiency and effectiveness of development expenditures first at the project level and later at their entire sectoral development program. Fourthly, both the central and the local governments, have increased their fiscal deficits making enormous long-term investments for physical and social infrastructure development such as power, transportation, communication, education, training, health and research and development (R & D). Increasing fiscal deficits have, in turn, demanded a greater efficiency and effectiveness of public development and routine expenditures. Finally, but no less important, in order to maximize their developmental and distributional impacts, bilateral and multilateral donors have been increasing pressure on their aid-recipient partner countries to first promote evaluation of their ODA projects and programs and then promote evaluation of all national development plans, programs and projects financed by their governments.

It must be remembered, however, that it is one thing for any developing country to realize the need for installing national evaluation machinery or even a national evaluation society, but it is quite another to actually install them. Furthermore, once established, it is quite an accomplishment to be able to maintain and manage a well-functioning and effective system of evaluation that can contribute to sustained economic and social development.

4) Major Issues of Evaluation Capacity Development in the Region

Reviewing specifically each of these Asia-Pacific countries that have gone ahead in strengthening national evaluation machinery and even in establish-

ing national evaluation societies, there have been in these countries distinct features of evaluation machinery and common thrusts of emphasis in evaluation programs, as well as challenges specific to their countries and organizations. Elsewhere national evaluation machinery, though established formally, is still being developed at different stages with the assistance of bilateral and multilateral donors.

a) National Evaluation Machinery

National evaluation machinery in the more advanced (evaluation-wise) developing countries of the region is designed and supervised by the Ministry of Planning (Bangladesh), the Ministry of Finance (India), Economic Planning Unit, Office of the Prime Minister and the Ministry of Finance (Malaysia), National Economic Development Authority (the Philippines), the Ministry of Plan Implementation (Sri Lanka), Economic Development Board (Singapore), or its equivalent organization such as the Ministry of Science and Technology (China). In these countries, the heads of these ministries are members of the Cabinet, with each of the Sectoral Ministries reporting their sectoral evaluation results regularly, i.e., quarterly or bi-annually, to the supervising ministry in accordance with the evaluation and reporting guidelines and procedures defined in National Assembly legislations whose details are defined by Cabinet directives set by the Prime Minister. Within each sectoral Ministry most governments have set up an evaluation division/unit charged with laying down specific evaluation procedures including evaluation manuals and supervising the sectoral evaluation. The evaluation unit within each sectoral Ministry has to report every year to the Minister the results of their evaluation.¹²

At the local government level, however, evaluation machinery has not been set up in most developing countries even in the Asia-Pacific region. This makes it more difficult for national evaluation machinery to function well in accordance with its objectives of improving development management for results for the whole country and increasing transparency and accountability to the people and other various stakeholders within and outside their own countries. The absence of local evaluation machinery in provinces, cities and villages is both a cause and effect of the inadequacy or complete lack of an evaluation culture in many developing countries. It is reasonable to conclude that in spite of the recent emphasis, in many developing countries, on decentralization of government administration the traditional political and administration setup makes it difficult, in terms of manpower, budget and

system, to strengthen the evaluation machinery of local provinces before strengthening the national evaluation machinery.

At both the national and local level, evaluation machinery, according to the OECD Survey of Agencies' Country Level Activities on Environment and Development, most developing countries in the region have had assistance, either bilateral or multilateral or both in various sectors in response to the Paris Declaration.¹³ Cambodia, Pakistan, India, Vietnam, and Mongolia to name a few. Since 2006, Cambodia has been assisted by the UNDP, with the involvement of New Zealand and U.K. aid agencies, in the development of national leadership and capacity in aid coordination as part of the coordinated external aid flow mechanism to be established by 2010 as a prerequisite for aid effectiveness enhancement. The UNDP Regional Centre in Bangkok, is providing on-going support to Pakistan focusing on practical advice on how to establish an effective coordination mechanism through the packaged capacity development measures and policy advice with customized aid and budgeting tracking technology. Pakistan has also received technical assistance from the UNDP to reinforce existing aid coordination capacities within the Economic Affairs Division for improved information outreach, analysis, database management and communication (all the most deficient areas essential to effective evaluation machinery in the country). With respect to India and Vietnam, the Coordination Unit through the Community of Practitioners now operating in the Asia-Pacific region aims to establish a Community of Practice on aid coordination both to link aid coordination practitioners from the Asia-Pacific region closer together and facilitate knowledge sharing and mutual support among interested countries of the region. Mongolia has also received UNDP support to develop national capacity in terms of technical skills and institutional reforms in aid coordination and management in the context of national programming and budgeting.

b) Major Issues of Monitoring and Evaluation Programs

The common thrusts of development evaluation programs in these countries consist mainly of project and program (sub-sector) evaluation and very rarely of policy evaluation. Challenges of implementing project and program evaluation are varied among these Asia-Pacific countries, but with the possible exception of India, Malaysia, the Philippines, Singapore and Sri Lanka, most outstanding are the enormous difficulty in getting the on time information and quantitative data required for quality evaluation at the regularly scheduled intervals, i.e., monitoring, as well as the timely and detailed analysis of

the monitored information for evaluation reporting to the supervisory unit of each Ministry. Essentially three factors are to blame for the difficulties, i.e., i) inadequate staffing of qualified evaluators within government ministries (supposedly due to financial constraints), ii) inadequate availability of domestic evaluation professionals within respective countries (more importantly), and iii) the lack of evaluation culture, concern and leadership within each Ministry, as mentioned earlier (most significantly).

India, Malaysia, the Philippines, Singapore and Sri Lanka have made remarkable progress in monitoring and evaluation programs due to strong national government leadership in implementing economic and social development programs effectively and efficiently. Malaysia and Singapore employed top down political leadership by officially adopting the “Look East Policy.” On the basis of strong support of the United Malay National Organization (UMNO) and the People’s Action Party (PAP), respectively the Look East Policy relied on former Malaysian Prime Minister Mahathir and former Singaporean Prime Minister Lee Kwan Yew to personally and vigorously pursue a broad-based and people-oriented development.¹⁴ Singapore, a small island republic after its separation from Malaysia in 1965, has under Mr. Lee Kwan Yew consistently pursued results emphasizing clean government, bureaucracy and management requiring the closest possible monitoring and evaluation of all government activities across all Ministries, Agencies and Public Institutions including Utility Boards. India, the Philippines and Sri Lanka, on the other hand, followed a bottom-up approach, continuing the decentralization of administrative authority in response to the rising demand from active civil society and traditionally strong local governments under political pluralism for giving priority to effective and equitable development at the local level. Closer to the people living in local communities, local governments have been subjected over the years to the closer scrutiny by local communities of all their activities, often supported and even precipitated by relatively free and independent mass media.

India has traditionally had a federal system of political and administrative setup where State Governments have enjoyed far more authority and stronger governance structure at the local/provincial level, as compared with many other developing countries in the region such as Cambodia, Laos, Myanmar, Thailand and Vietnam. Opposition parties in India, as in the Philippines and Sri Lanka, have been strong and rather effective in insisting on the public scrutiny of government activities and budget expenditures. In some countries, as in Bangladesh where national and even local monitoring

and evaluation machinery had been fairly developed, political turmoil has in recent years affected the effective operation of the established evaluation machinery at the national and local levels. With somewhat successful general election completed and the installment of a new government in December, 2008, it is the sincere wish of everyone concerned with evaluation capacity building in Asia and in particular of Bangladesh, that Bangladesh national and local evaluation machinery and national evaluation society will resume its traditionally active engagements and activities in evaluation and join other national evaluation societies in Asia in pursuing evaluation capacity development in the region.

In most developing countries of the region, however, evaluation is rarely integrated into policy review and policy making within each Ministry, because Ministers tend to prefer their own “new” policy and program initiatives based on their personal observations and interests rather than project and program evaluation results already carried out in each Ministry. These bottlenecks of evaluation implementation suggest that not only does the understaffing and under-financing of the evaluation division/unit within each Ministry have to be rectified, but also evaluation culture has to be developed and steadily nurtured at all ranks of the bureaucracy within each Ministry in order for these countries to implement effective Monitoring and Evaluation of all their development projects and programs.

c) National Evaluation Societies

In the developing Asia-Pacific region, a national evaluation society has existed for some time in Bangladesh, India, Malaysia, Pakistan and Sri Lanka. Nepal founded its national evaluation society in March, 2009. In the developed A/P region, Australia, Japan and the Republic of South Korea have had such society for some time. Some of them are strong and active, while others are not so. It is reported that in Thailand and Vietnam, national evaluation societies are now in the process of organization. It is likely that Cambodia, Indonesia, Laos, Mongolia and the Philippines will be next in line for establishing a national evaluation society. China has a strong national evaluation machinery established but does not yet have an evaluation society. It is also understood that Myanmar does not yet have an evaluation society. There are, however, national government agencies constituting the Community of Practitioners in the Asia-Pacific region, sharing knowledge and experiences in all aspects of aid management and coordination in their respective countries.

In all those countries where national evaluation societies exist, they are registered with their respective government authorities, specifically the ministry of the interior, or the ministry of local government, or the ministry of community development. In some countries bureaucratic red tape tends to hinder fast and easy registration of societies, and in some other countries registration of societies tends to be restricted on political grounds. Government authorities require a national evaluation society to register its organizational name, office address, constitution and by-laws, elected officers and decision-making bodies. Also, every country has mandatory requirements for registered national evaluation society to have its annual general meetings (AGM) where final decisions are made on any proposals made by their board of directors (management) in regards to constitutional changes, program, membership composition, operational and procedural modalities including membership fees, etc. At the AGM, the highest decision-making organ of a national evaluation society, which is held at the end of the fiscal year, the minutes of the preceding board meetings during the current year are confirmed, which usually contain the results of any elections made, the business performance, the decision on the accounts and auditing of the society for the current year and its proposed program and budgets during the coming year(s). And when requested, any other matters brought up by its membership to the AGM are also discussed and decided on. Unfortunately, however, in some countries (without naming which ones) national evaluation societies do not have enough members, revenues and business activities and program and active participation in the debates at such AGM. Governments of these countries do not seem to rigorously enforce their reporting requirements.

Where relatively strong and active, national evaluation societies (Bangladesh, India, Malaysia and Sri Lanka) are most often composed of serious scholars and experienced evaluation practitioners in public and private sector organizations. They are engaged in updating the current evaluation programs, practices and information through regular newsletters for the benefits of their members and promoting the exchange of evaluation experiences at home and overseas among their members and with outside organizations. These societies are also encouraging their members through journal/periodical publication to come up and propagate with innovations in evaluation concepts, methodologies, programs and systems at sectoral, regional (sub-national) and national levels. They provide a monthly forum where guest speakers are invited to speak on selected topics related to evaluation.

They also conduct seminars and workshops to train and update potential and current evaluation practitioners on evaluation skills, procedures and methodologies.

These national evaluation societies also often send their delegations to regional and international evaluation conferences and symposiums overseas for further exchange of their evaluation ideas and experiences with their counterparts. In addition, their programs include joint evaluation studies often initiated by their members in partnership with their counterparts in other national evaluation societies abroad, as well as the provision of consulting services for local and national governments, bilateral donors and international organizations. In fact, joint evaluation by concerned agencies of donor and partner countries has been observed mainly in those countries where there is a national evaluation society with active evaluation professionals and experiences. They often collaborate with OECD/DAC Working Party on Evaluation and such international evaluation societies as IDEAS and IOCE to promote strategic alliances to facilitate the formulation of international guidelines on the standardization of evaluation procedures and formats.

Although a country's national evaluation society may be active in evaluation of policy, program and project formulation and implementation, it can still suffer from a lack of active members. In this case, the day-to-day operating burdens tend to fall on the board of directors and a few active volunteer members and the society may find itself always struggling with a shortage of financing for their officially announced activities, unless they have sponsoring organizations such as government ministries and bilateral and multilateral donors. Unfortunately, there are a few national evaluation societies in the Asia-Pacific region that are not as strong and as active as those mentioned above. It is fair to say, however, that they are trying hard to develop public awareness and appreciation of evaluation through newsletters and in their own way to expand their membership and provide services similar to those offered by more advanced organizations by organizing seminars and workshops with the help of visiting foreign evaluation experts.

d) Pre-conditions for Installing National Evaluation Societies

In the Asia-Pacific region although some countries do not have a national evaluation society, they do in fact have national evaluation machinery already installed. These countries include Bhutan, Cambodia, China, Indonesia, Lao, Papua New Guinea, the Philippines, Singapore, Solomon Islands, Thailand and Vietnam. There are a variety of reasons why the formal and/or substan-

tial installation of national evaluation machinery and a national evaluation society have, so far, not gone hand in hand. Without going into the specific details of each country, the following seem to be some of the more plausible reasons.¹⁵

First, as often said before, evaluation culture is not permeating throughout the different strata of the population and society. Installing national evaluation machinery is considered to be part of the government's agenda and not part of civil society's agenda. Second, related to the first reason, civil society has not taken action and NGOs as well as the public and private sectors remain unaware of corporate social responsibility, resulting in the lack of interest among various stakeholders of setting up a national evaluation society. Third, national and local governments may still have lingering suspicions that the conduct of national evaluation society threatens or challenges governmental authority over the formulating, implementing and managing of development policies, programs and projects. Evaluation, particularly evaluation of development policies and programs, is still considered to be a prerogative of the national and local governments and/or the ruling political party, and not to be devolved to any other stakeholders including the organization of evaluation professionals and practitioners. Finally, but not less important, enough local evaluation professionals and experts do not exist in the countries to get together and establish a forum including national evaluation society.

In spite of such welcome developments including greater evaluation awareness and improved national evaluation machinery in many more countries of the Asia-Pacific region, it should be admitted that it is not easy for many other developing countries in this region to develop all these technical, professional and managerial capacities required for high-quality and effective development evaluation within a short period of time, let alone to establish a national evaluation society. It is therefore vital that developed countries associated with the Organization for Economic Cooperation and Development (OECD), including Japan, as well as multilateral donors assist these developing countries to develop their evaluation systems at the local and central government levels. In this connection, it must be emphasized that the ODA Evaluation Workshop organized by the Ministry of Foreign Affairs, Government of Japan every year since 2000 and those Evaluation Seminars organized since some time ago by the former Japan International Cooperation Agency (JICA) and the former Japan Bank for International Cooperation (JBIC), now joined together into the new JICA, has contributed

enormously to developing evaluation culture as well as evaluation capacity building in participating countries.¹⁶ It should be hastened to add that the Japan Evaluation Society has had some positive role in this process of evaluation capacity development in several developing countries in the Asia-Pacific region through technical cooperation on evaluation methodologies, practices and machinery.¹⁷

4. Recent Developments of Regional and International Evaluation Associations

As corporate and government evaluation activities have become prevalent in every sector and multinational corporation in every region of the world, and as national evaluation societies expand and deepen their professional activities in response to the changing needs and requirements of the corporate world and the general public at home and overseas, the need for exchanging evaluation information and experiences among countries has become greater and sharply focused. This need was first felt most strongly among national evaluation societies in developed countries, resulting in the establishment of the European Evaluation Society¹⁸ in 1994, the Australasian Evaluation Association¹⁹ in 1997 and the African Evaluation Association (AfrEA)²⁰ in 1999. While the Australasian Evaluation Association has a regional coverage, its activities are focused on Australia, New Zealand and some South Pacific countries. Also, while there is no regional evaluation association (REA) in North America, the American Evaluation Association and the Canadian Evaluation coordinate and collaborate in terms of their programs such as coordinating the core themes and timing of their respective annual conventions. Unfortunately, active REAs covering Asia-Pacific, Latin American and the Caribbean as well as the Middle East do not exist. However, at an inter-governmental level, the Community of Practitioners in the Asia-Pacific region and elsewhere share experiences on all aspects of aid management and coordination. Whereas AfrEA has been established and has organized biennial conventions for some time, it has depended too much on the initiatives of the South African Evaluation Association headquartered in Johannesburg. Together with OECD/DAC Working Party on Evaluation, bilateral and multi-lateral donors have been assisting developing countries strengthen evaluation capacity and also participate in regional REAs, to learn lessons from the activities of other national evaluation societies.

Although employing quite different financing practices, REAs essentially

share similar objectives, programs, modality of operation and organizational structure. Dual objectives are common in all REAs. The first is to promote evaluation culture and improve evaluation methodologies and practices in member countries. The second is to promote the exchange of evaluation information and experiences among member countries. In recent years, with the establishment of the International Development Evaluation Associations (IDEAS) in 2002 and the International Organization for Cooperation in Evaluation (IOCE) in 2003, REAs increasingly link their activities with international evaluation associations with a view to presenting a common stand vis-à-vis member countries' governments as well as bilateral and multilateral donors. They take a common stand in regards to the need for setting up independent evaluation programs as well as encouraging greater participation of civil society in all corporate and government policy, program and project evaluations.

As far as programs are concerned, all REAs are concerned with the need for reviewing evaluation policies and practices in member countries. In this regard, the long-held OECD practice of PEER REVIEW is considered an appropriate approach to evaluation on all levels: project, sector or country. Also, REAs are interested in pooling their technical and organizational expertise to support evaluation capacity building in member countries considered to have inadequate capacity. Training of evaluation professionals in member countries is of course the responsibility of national evaluation societies, but becomes REAs responsibility in those member countries where there is no national evaluation society or training institutions. Often, in this case, REAs receive assistance from international aid agencies and other national or international evaluation associations. REAs are membership-based organizations, with their board of directors elected by their members, individual and corporate, at their regular general assembly for a specified term, e.g., two to three years, and their executive/management committees are elected by the board for the same term length. The annual or biennial general assembly is the final decision-making machinery of all REAs, but the day-to-day operation of the associations is left to the executive/management committee under the supervision of the board of directors. Financing for REAs comes from annual membership fees and contributions by interested parties such as bilateral and multilateral donors, and public and private foundations either on the basis of study/survey assignments or free-standing donations.

It is interesting to observe that in recent years, as evident by the top agenda item for REAs and national evaluation societies such as the conven-

tions organized by both the European Evaluation Society's annual convention in Lisbon in October 2008 and the Japan Evaluation Society at its 9th annual convention in Kyoto in November 2008, that REAs are increasingly concerned with the over-riding question of the usefulness of evaluation at all levels and the utilization of the evaluation results and reports by policy-makers. Their common concern stems from the fact that in spite the professional evaluation community's increased emphasis on the "Outcome" rather than the "Input" and the "Output," outcome-oriented evaluation has not become as pervasive as expected in both developed and developing countries. As a result, there seems to be a continued under-utilization of evaluation results by policy-makers, keenly interested in policy and program effectiveness both in the bureaucracy and the parliament, as well as by the relevant civil society organizations, (including non-governmental organizations (NGOs)) keenly concerned with the impact on the intended beneficiaries and the rest of the population.

Under-utilization also stems from the fact that in spite of the increased interests among civil society and the general public in strategic evaluation, both the government and professional evaluation community have been overly concerned with the comprehensiveness of evaluation regarding the stated multiple objectives of evaluated projects, programs and policies, thus reducing the usefulness of strategic evaluation to the community and not meeting the expectations for evaluation among the masses of people. Civil society participation in evaluation is essential not only to promote evaluation culture, enrich evaluation experiences and develop evaluation expertise, but also to make evaluation usable by and useful to the general public. In view of the fact that development evaluation has today become too complex and time-consuming for many developing countries equipped with smaller evaluation capacity, efforts will have to be made in the future for both national evaluation societies and REAs to work together and jointly propose a much simpler approach to out-come-based evaluation which, for example, requires far less documentation for data collection and analysis.

5. Concluding Remarks and Proposal for Establishing Asia-Pacific Evaluation Association Network (APEA NET)

For these reasons mentioned above, the author strongly recommends an early establishment of the Asia-Pacific Evaluation Association Network. At

the recent meeting of the Japan Evaluation Society held on 30th November 2008 at Doshisha University in Kyoto, the International Affairs Committee of the Japan Evaluation Society circulated a preliminary draft proposal (based on the earlier endorsement of the proposed establishment of the Asia-Pacific Evaluation Association by the Asian Regional Forum on Aid Effectiveness held in October 2006 in Manila) to establish a preparatory committee for the establishment of APEA NET.²¹ Participants from Japan, Nepal and Vietnam agreed that a formal draft be presented for consideration and adoption at a special session following the 8th ODA Evaluation Workshop to be held in Singapore in March 2009. It is envisaged that APEA NET would be a non-profit, non-government, volunteer organization with membership open to individuals and institutions involved in evaluation as well as in development from developing and developed countries. Furthermore, members of APEA NET would include evaluation experts and development practitioners from governments, civil society (e.g., NGOs, academia, research institutions), the private sector, and bilateral and multilateral institutions in the international development cooperation community. Interested parties in Asia-Pacific region are expected to finalize a formal draft sometime in 2009 for the establishment of a preparatory committee for the founding of APEA NET possibly later in 2010.

1) Objectives and Missions

The main objectives of the proposed APEA NET are as follows. First, to promote theories, practices and utilization of evaluation, in particular a quantitative approach to and methodologies of results-based evaluation and process evaluation to ensure professionalism, objectivity, neutrality/independence and a participatory approach through joint studies, seminars and conferences in Asian and Pacific countries. Second, to enhance academic and professional networking among evaluators and others concerned with evaluation in the region through an exchange of evaluation information and experiences, in particular through participation in peer review of aid policies and practices of donor countries and multilateral aid institutions and active participation in joint evaluation exercises conducted either by or for bilateral and multilateral aid agencies. And third to assist developing countries to enhance their evaluation capacity, both human and institutional, including, if possible, the setting up of an independent national evaluation committee or agency such as the Government Accountability Office of the U.S. Congress or the installation of an independent national evaluation fund with its own budgetary allocation

outside the executive branch as well as the initiation of national evaluation societies where none exist. In summary, it is expected that the missions of APEA NET would be to: promote a culture for evaluation; develop and improve evaluation expertise; and strengthen the evaluation capacities in all its member countries with a view to conducting high-quality evaluation researches and evaluation practices, and thus ultimately contributing to more efficient and effective economic and social development of its member countries in the region. APEA NET is also expected to cooperate with both regional evaluation associations such as the European Evaluation Society and international associations such as IDEAS and IOCE in pursuit of the above objectives.

2) Organization and Management of the Proposed APEA NET

As tentatively agreed upon at the Kyoto meeting, APEA NET is envisaged to have the following management structures, subject to further amendments.

- i) APEA NET General Assembly (GA): the highest decision-making organization represented by all institutional and individual members and meeting once a year for all the major decisions of APEA NET.
- ii) APEA NET Board of Management (BOM): the policy-formulating body elected by APEA NET General Assembly and accountable to it.
- iii) APEA NET/BOM shall have a few committees specifically responsible for proposing annual and longer-term business plans and budgets, membership communications including newsletters as well as organizational management matters such as election of BOM and Executive Committee members.
- iv) APEA NET/Executive Committee (EC): the policy implementation body consisting of President, Vice President, Treasurer and Secretary.
- v) APEA NET/Secretariat: Secretary heads a small secretariat in charge of running the day-to-day business and communications of APEA NET internally and externally on EC's behalf, and to minimize the cost of operation, the Secretariat should be located at one of the national evaluation societies associated with APEA NET, with separate and independent accounts registered and maintained for APEA NET.
- vi) Any other matters.

3) Financing of the Proposed APEA NET Activities

There may be many approaches to financing APEA NET once it is established. After its formal launching sometime in late 2009, the following financing schemes can be considered.

- i) Annual membership fees: institutions US\$X for LICs, US\$X times 1.5 for MICs, and US\$3X for HICs (Australia, Japan, New Zealand, Republic of Korea, Singapore and Taipei, China); individuals US\$X times 0.2 from developing countries (LICs and MICs) and US\$X times 0.5 from high-income countries (HICs).
- ii) Contributions and donations.
- iii) Contract funding.

Appendix

Table 1. Millennium Development Goals So Far Achieved in All Regions

Regions	A	B	C	D	E	F	G
World	23.9	86	95	72	65	82	1.0
East Asia & Pacific	13.3	98	100	29	87	91	0.2
Europe & Central Asia	n.a.	98	97	26	95	97	0.6
Latin America & Carib.	5.1	100	101	26	88	90	0.6
Middle East & N. Africa	n.a.	90	93	42	77	73	0.1
South Asia	41.3	80	90	83	41	58	0.7
Sub-Saharan Africa	26.8	60	87	157	45	59	5.8
High income	n.a.	97	99	7	99	99	0.4

Source: World Bank, World Development Report (WDR) 2009.

Notes: A stands for prevalence of child malnutrition % of children under 5, 2000-07; B for primary education completion rate 2006; C for ratio of girls to boys enrollments in primary and secondary school % 2006; D for under 5 mortality rate per 1000, 2006; E for births attended by skilled health staff % of total 2000-07; F for adult literacy rate % ages 15 & over 2005; and G for HIV prevalence % of population ages 15-49, 2005.

Table 2. Socio-Economic Indicators of Development in All Regions

Regions	A	B	C	D	E	F	G	H
World	1.2	28	7,958	9,852	3.2	40.8	n.a.	4.3
East Asia & Pacific	0.8	23	2,180	4,937	8.9	46.9	38.7	3.3
Europe & Central Asia	0.0	19	6,051	11,115	6.1	28.1	8.9	7.1
Latin America & Carib.	1.3	29	5,540	9,321	3.6	57.0	16.6	2.6
Middle East & N. Africa	1.8	32	2,794	7,385	4.5	34.4	16.9	3.8
South Asia	1.6	33	880	2,537	7.3	36.8	73.5	1.1
Sub-Saharan Africa	2.5	43	952	1,870	5.0	43.7	73.0	0.9
High income	0.7	18	37,566	36,100	2.4	24.9	n.a.	13.1

Source: UNDP, Human Development Report 2007/08; World Bank, *ibid.* and World Bank, Global Economic Prospects 2009.

Notes: A stands for Average annual % growth of population 2000-07; B for Population age composition % ages 0-14, 2007; C for per capita GNI 2007; D for per capita PPP GNI 2007; E for average annual % growth of gross domestic product (GDP), 2000-07; F for Gini index with World represented by a proxy of the United States, the worst among high income countries, High Income group by Japan, the best among the group, and all the other country groups by proxies of the most populous country in each group; G for population below US\$2 a day as percent of the total population, 2005; and H for carbon dioxide emission per capita in metric tons 2004.

Table 3. Unemployment, Income and Gender Inequality and Access to Electricity in Asia-Pacific Countries

Countries	A	B	C	D	E	F	G
Singapore	5.3	98	1.9	42.5	0.0	0.92	0.51
Republic of Korea	4.0	76	2.9	31.6	n.a.	n.a.	0.40
Malaysia	3.6	100	1.7	49.2	0.6	0.93	0.36
Thailand	1.4	80	2.7	42.0	0.6	0.95	0.62
China	4.2	n.a.	1.6	46.9	8.5	0.91	0.64
Philippines	7.4	99	2.2	44.5	16.2	1.02	0.61
Sri Lanka	7.7	216	3.0	40.2	6.7	0.97	0.41
Vietnam	2.1	131	4.2	34.4	13.2	0.93	0.70
Indonesia	9.1	155	3.6	34.3	101.2	0.92	0.46
Mongolia	3.3	120	3.0	32.8	1.0	1.00	0.50
India	4.3	100	3.6	36.8	487.2	0.65	0.31
Lao, PDR	n.a.	n.a.	3.4	34.6	n.a.	0.79	0.51
Cambodia	1.8	147	2.9	41.7	10.9	0.76	0.74
Myanmar	n.a.	n.a.	n.a.	n.a.	45.1	0.92	n.a.
Pakistan	7.7	194	4.0	30.6	71.1	0.55	0.29
Bangladesh	4.3	117	3.7	33.4	96.2	0.76	0.46
Nepal	1.8	85	2.6	47.2	18.1	0.56	0.50
Papua New Guinea	2.8	30	1.7	50.9	n.a.	0.80	0.72

Source: UNDP, HDR 2007/2008.

Notes 1: A stands for unemployment rate 1996-2005; B for female unemployment rate as % of male rate 1996-2005; C for share of income/consumption of the poorest 10% of the population for different years in the years 1998-2004; D for Gini index; E for population without electricity in millions 2005; F for the ratio of adult female literacy rate to male 1995-2005; and G for the estimated earned income of female to male 2005.

Note 2: Regarding Republic of Korea, reference period of data in A and B are 1996-2006.

Table 4. Investment, Trade, Aid and Finance in All Regions

Regions	A	B	C	D	E	F	G1	G2
World	22	0	0.2	1.9	5.1	n.a.	n.a.	n.a.
East Asia & Pacific	38	7	0.2	n.a.	n.a.	n.a.	-1.1	-0.6
Europe & Central Asia	24	-1	0.3	n.a.	n.a.	n.a.	2.6	2.9
Latin America & Carib.	22	1	0.3	2.9	6.6	22.9	1.2	1.4
Middle East & N. Africa	26	-1	3.0	n.a.	n.a.	n.a.	5.5	0.7
South Asia	35	-4	0.8	0.8	2.6	15.4	-5.9	-6.1
Sub-Saharan Africa	21	-3	5.1	2.4	n.a.	n.a.	0.2	1.0

Source: World Bank, *ibid.*; for A and B; and United Nations Development Programme (UNDP), Human Development Report (HDR) 2007/2008 for C, D, E, F and G.

Notes: A stands for gross capital formation as % of GDP 2007; B for external balance of goods and services as % of GDP 2007; C for official development assistance as % of GDP 2005; D for net foreign direct investment inflows as % of GDP 2005; E for total debt service payments as % of GDP 2005; F for total debt service payments as % of the exports of goods, services and net income from abroad 2005; and G1 and G2 for fiscal balance as % of GDP respectively in 2005 and 2006.

Table 5. Governance in Asia-Pacific Countries, 2000

Countries	A	B	C	D	E	F	G	H
	-10~10	7~1	7~1	0~100	2.50~ -2.50	2.50~ -2.50	2.50~ -2.50	2.50~ -2.50
Singapore	-2	5	5	68	0.11	1.85	2.16	2.13
Brunei D.	n.a.	5	7	74	-0.93	1.29	0.88	-0.17
Malaysia	3	5	5	70	-0.13	0.34	0.53	0.13
Thailand	9	3	2	29	0.37	0.44	0.10	-0.46
Philippines	8	3	2	30	0.53	-0.49	0.03	-0.49
Vietnam	-7	6	7	80	-1.29	-0.57	-0.30	-0.76
Indonesia	7	4	3	47	-0.40	-0.87	-0.50	-1.01
Myanmar	-7	7	7	100	-1.93	-1.02	-1.25	-1.18
Cambodia	2	6	6	61	-0.77	-0.38	0.34	0.34
Lao, PDR	-7	6	7	69	-1.05	-0.72	-0.39	-0.31
China	-7	6	7	80	-1.11	-0.19	0.14	-0.30

Source: UNDP, HDR 2002, Table A1.1, pp.38-41.

Notes: A stands for "polity score" ranging from -10 (authoritarian) to 10 (democratic); B for "civil liberties" ranging from 1 (free) to 7 (not free); C for "political rights" ranging from 1 (free) to 7 (not free); D for "press freedom" ranging from 0 (having a free press) to 100 (having a press that is not free); E for "voice and accountability" ranging from -2.50 to 2.50 (higher is better); F for "rule of law" ranging from -2.50 to 2.50 (higher is better); G for "government effectiveness" ranging from -2.50 to 2.50 (higher is better); and H for Graft within "corruption" ranging from -2.50 to 2.50 (higher is better).

Note 2: Data in E, F, G, and H range from 2000 to 2001.

Table 6. Governance Indicators in Asia-Pacific Region, 2007

Regions	Percentile Rank (0~100)	Governance Score (-2.5~+2.5)
East Asia and the Pacific		
Voice and Accountability	49.4	-0.03
Political Stability	59.7	+0.34
Government Effectiveness	48.1	-0.09
Regulatory Quality	45.2	-0.18
Rule of Law	54.9	+0.15
Control of Corruption	45.2	-0.19
South Asia		
Voice and Accountability	28.0	-0.69
Political Stability	19.3	-1.32
Government Effectiveness	36.5	-0.50
Regulatory Quality	31.9	-0.61
Rule of Law	38.3	-0.46
Control of Corruption	33.9	-0.56

Source: World Bank, Worldwide Governance Indicators 2007

Note: Governance score ranges from about -2.5 to 2.5, with higher values corresponding to better governance outcomes.

Notes

1. See the IDEAS website. The International Development Evaluation Associations (IDEAS) was launched with its inaugural convention held in Beijing in 2002 which was participated by more than 100 individual experts and national and international organization representatives, with a view to promoting development evaluation in all countries through studies, seminars, workshops and conferences organized by IDEAS in collaboration with national and regional evaluation societies, bilateral and multilateral donors and developing partner countries. Initially its focused programs were: a) rethinking development evaluation, b) governance and accountability for development, and c) strengthening poverty-environment nexus. Since then, IDEAS has expanded its program by reaching out to national development evaluation networks in developing countries as part of its overall assistance program priority. (www.IDEAS-int.org)
2. See the IOCE website. It took some time after formal launching of the IOCE in Lima in 2003 that IOCE was able to organize its priority program due to the lack of substantive and financial support by potential members and related organizations. The organization's main objectives based on institutional membership are to promote exchange of information and experiences related to all aspects of evaluation, unlike IDEAS which confines its activities related to development evaluation and which accepts individual and institutional experts as its full members. It is especially concerned with establishing and propagating ethics and standards of evaluation and training of evaluation personnel to achieve high quality evaluation in both private and public sector organizations. (www.IOCE-int.org)
3. Japan Evaluation Society (JES), established in 2000 and with its membership close to 500 as of November 2008, has programs similar to other national evaluation societies, such as study conventions, publication of Japanese- and English-language journals and newsletters, consulting services, membership drive, presentation of JES awards to outstanding members both in terms of papers presented and contribution to JES activity and maintenance of its official website in Japanese language. JES has so far conducted a total of 9 annual national conventions (autumn) and 5 spring conventions where numerous papers have been presented by its members and non-members on their findings on policy, program and project evaluation in all sectors at home and overseas. JES has begun

since one year ago a certification program for junior evaluation specialists. JES has been a co-organizer of the ODA Evaluation Workshop sponsored by the Ministry of Foreign Affairs, the Government of Japan every year since 2001 where most of Asian and Pacific governments have sent their delegations from their respective ministries and agencies charged with evaluation in the public sector. The author, the then vice president of JES, founding member of IDEAS established in 2002, served as vice president of IDEAS for two consecutive terms during the initial years 2002-2006 and many JES members have participated and presented their study findings at IDEAS' conventions and other international conferences at home and abroad. (www.idcj.or.jp/JES/)

4. Sri Lanka Evaluation Association (SLEvA), established in 1999 and with current membership just over 80, is composed of those interested in evaluation across all sections of Sri Lankan society, academia, and professionals and managers in public and private sector organizations. Its program resembles those of other national evaluation societies, conducting seminars, training and conferences where papers are presented by its members and non-members on their research findings. They also provide consulting services and their members participate in conferences organized by international and regional evaluation societies and bilateral and multilateral organizations. SLEvA has had a strong support from the Ministry of Plan Implementation and Monitoring and Evaluation Agency of the Sri Lankan Government, as well as from international organizations such as United Nations Development Programme and UNICEF. (www.nsf.ac.lk/sleva/)
5. Malaysian Evaluation Society (MES), officially registered in 1999 and with membership nearly 100, has its membership from all sections of Malaysian society, academia and those interested in evaluation both in private and public sectors. Its aim is similar to that of other national evaluation societies, i.e., promoting the ideas, awareness and studies on evaluation throughout the Malaysian society by way of organizing studies, seminars and workshops at national and local levels as well as through its newsletters, periodicals and website. It has so far successfully organized three international conferences with the participation of practitioners, academia and other professionals in Malaysia, regional countries and bilateral and international organizations. The third international conference in March 2008 was organized by MES and IDEAS. (www.mes.org.my)

6. See *OECD/DAC* (2006), Report and Proceedings of the 2006 Asian Regional Forum on Aid Effectiveness: Implementation, Monitoring and Evaluation. (www.1.oecd.org/dac/evaluation/)
7. Currently all OECD countries, with the exception of the Republic of Korea and Mexico, have national evaluation societies whose aims are: a) to advance theories and practices and utilization of evaluation, b) to provide a forum where evaluation academics, practitioners and other experts are able to present their thinking and publish articles on all aspects of evaluation, c) train evaluators to meet the changing needs and requirements of organizations and larger society, and d) to assist evaluation capacity building in developing countries. (See American Evaluation Association (www.eval.org), Canadian Evaluation Association (www.evaluationcanada.ca), French Evaluation Society (www.sfe.asso.fr), German Evaluation Society (www.degeval.de), Italian Evaluation Society (www.valutazione.it) and United Kingdom Evaluation Society (www.evaluation.org.uk).) An increasing number of developing countries have set up national evaluation societies under assistance by bilateral and multilateral donors such as United Nations, World Bank and regional development banks as well as by regional and international evaluation associations.
8. See *IMG, Inc.* (2004), A Survey on the Evaluation System of major Donors: English Summary.
9. (www.un.org/millenniumgoals)
10. (www.worldbank.org/gmr2008)
11. *European Commission* (2008), Millennium Development Goals at Midpoint, Brussels: EC.
12. See a series of the Reports and Proceedings of ODA Evaluation Workshops organized by the Government of Japan, Ministry of Foreign Affairs since 2001.
13. See *OECD* (2007), Survey of Agencies' Country Level Activities on Environment and Development, Paris: OECD.
14. See *Hirono, Ryokichi*, "Economic Development and Social Values in Singapore," *Seikei University Journal of Economics*, vol. 22, No. 1, 1986.
15. For a closer examination of some of these reasons behind the late development of national evaluation machinery and professional national evaluation society in some of these Asia-Pacific countries, see *Minato, Naonobu and Tadashi Kikuchi*. "Tojoukoku no Hyouka Nouryoku no Koujo: Betonamu no Jirei (Improvement of Evaluation Capacity in

- Developing Countries: Case of Vietnam), in *Minato, Naonobu* (ed.) (2008), *Kokusai Kaihatsu ni okeru Hyouka no Kadai to Tenbo* (Major Issues and Prospects of Evaluation in International Development), Tokyo: FASID, Chapter 8, pp. 117-141; *Dhaerani Dhar Khatiwada and Subaruna Lal Shrestha* (2008), Progress and Achievements of the Nepalese Government's Efforts to Improve the Evaluation System, presented at the 9th Annual Convention of Japan Evaluation Society at Doshisha University, Kyoto in November 2008, and *Cao Manh Cuong* (2008), Evaluation Capacity Development in Vietnam, presented at the above Annual Convention.
16. See as examples of such workshops organized by the Japanese Government, *Ministry of Foreign Affairs* (2007), Report and Proceedings of the Workshop on ODA Evaluation in Thailand and MOFA (2008), Report and Proceedings of the Workshop on ODA Evaluation in Malaysia, Tokyo: MOFA. Also, for Japanese Government assistance to Nepal for strengthening the monitoring and evaluation system in Nepal, *Ishida, Yoko, Junko Miura and Yoko Komatsubara* (2008), Support for Institutional Strengthening of the Monitoring and Evaluation System in Nepal, presented to the 9th Annual Convention of Japan Evaluation Society held at Doshisha University, Kyoto on 30th November, 2008, and for the Japanese Government assistance to Vietnam for the same purpose, *Japan Bank for International Cooperation and Ministry of Planning and Investment, Government of the Socialist Republic of Vietnam* (2008), Implementation Program on Joint Evaluation Scheme 2008 during the Period from June 2008 to June 2009.
 17. See *Hirono, Ryokichi* (2007), Essentials of Evaluation, presented at the Joint Nepal-Japan Evaluation Forum at Yak & Yeti Hotel, Kathmandu, Nepal on 18 December, 2007, and *Hirono, Ryokichi* (2008), Japan Evaluation Society: Missions, Goals. Achievements, Programmes and Institutional Development, presented at the Feedback Workshop on Vietnam-Japan Joint Evaluation Program 2007, held in Hanoi on 23 June, 2008.
 18. See European Evaluation Society (EES) website. EES, founded in 1994 and starting its work in January, 1996, aims at promoting evaluation theory, practices and utilization especially in European countries, by bringing together academics and practitioners in all sectors from all over Europe to its meetings, workshops and annual conventions. (www.europeanevaluation.org)

19. See Australasian Evaluation Association (AEA)'s website. AEA, founded in 1997, is a non-profit, membership-based organization with over 1000 members and chapters in various key cities in Australia and New Zealand. Its main aims are establishing and propagating ethics and standards in evaluation practices as a service to community, advancing evaluation theories, practices and uses through regular publication of innovative evaluation articles in its journals and providing evaluation information through its newsletter. It provides education and training in evaluation and encourages networking among its members and those concerned with evaluation. (www.aes.asn.au)
20. See AfrEA's website. "AfrEA was founded in 1999 in response to a growing demand for information sharing, advocacy and advanced capacity building in evaluation in Africa. It is an umbrella organisation for national M&E associations and networks in Africa, and a resource for individuals in countries where national bodies do not exist. AfrEA works with the national networks and committed donors to develop a strong African evaluation community. It has held "four" African evaluation conferences, with a "fifth" to take place in 2009. It has facilitated the development of African Evaluation Guidelines to enhance the quality and utility of evaluation on the continent. A database of evaluators on the AfrEA website highlights African evaluation expertise. The Africa Gender and Development Evaluation Network, an AfrEA SIG in partnership with UNIFEM, has been enhancing African capacity in gender and rights-based evaluation since 2003. AfrEA continues to work with its member associations, interested organisations, donors and others to build capacity across the continent, establish networks that develop evaluation methods and share expertise, do research to improve evaluation practice and theory, and advocate the role African evaluators should play on the continent." (www.afrea.org)
21. See *Hirono, Ryokichi* (2008). A Preliminary Draft Proposal for the Establishment of a Preparatory Committee for the Formal Launching of the Asia-Pacific Evaluation Association Network (APEA NET), presented at the 9th Annual Conference of Japan Evaluation Society at Doshisha University, Kyoto on 30th November, 2008.

Institutionalizing Impact Evaluation Systems in Developing Countries: Challenges and Opportunities for ODA Agencies¹

Michael Bamberger

1. Introduction

With the growing emphasis on the assessment of aid effectiveness and the need to measure the results of development interventions, many development agencies now recognize that it is no longer sufficient to simply report how much money has been invested in their programs or what outputs (e.g., schools built, health workers trained) have been produced. Parliaments, finance ministries, funding agencies, and the general public are demanding to know how well development interventions achieved their intended objectives, how results compared with alternative uses of these scarce resources, and how effectively the programs contributed to broad development objectives such as the Millennium Development Goals and the eradication of world poverty.

These demands have led to an increase in the number and sophistication of impact evaluations (IE). For example, for a number of years the Foundation for Advanced Studies on International Development (FASID) has been offering training programs on impact evaluation of development assistance for official Japanese ODA agencies, and for NGOs and consulting firms involved in the assessment of official and private Japanese development assistance². In the

1. This chapter draws extensively on the 2009 publication prepared by the present author for the Independent Evaluation Group of the World Bank “*Institutionalizing Impact Evaluation within the Framework of a Monitoring and Evaluation System*”. The author would like to thank Nobuko Fujita (FASID) for her helpful comments on an earlier draft of this chapter.

*most favorable cases, impact evaluations have improved the efficiency and effectiveness of ongoing programs, helped formulate future policies, strengthened budget planning and financial management, and provided a more rigorous and transparent rationale for the continuation or termination of particular programs*³. However, many IEs have been selected in an ad hoc and opportunistic manner, with the selection often depending on the availability of funds or the interest of donors; and although they may have made important contributions to the program or policy being evaluated, their potential contribution to broader development strategies was often not fully achieved.

Many funding agencies and evaluation specialists have tended to assume that once a developing country government has seen the benefits of a few well-designed IEs, the process of building a systematic approach for identifying, implementing, and using evaluations at the sector and national levels will be relatively straightforward. However, many countries with decades of experience with project and program evaluation have made little progress toward institutionalizing the selection, design, and utilization of impact evaluations.

This chapter describes the progress being made in the transition from individual IE studies in developing countries to building a systematic approach to identifying, implementing, and using evaluations at the sector and national levels. The chapter also discusses the roles and responsibilities of ODA agencies and developing country governments in strengthening the institutionalization of IE. When this is achieved, the benefits of a regular program of IE as a tool for budgetary planning, policy formulation, management, and accountability begin to be appreciated. To date, the institutionalization of IE has only been achieved in a relatively small number of developing countries, mainly in Latin America; but many countries have already started or expressed interest in the process of institutionalization. This chapter reviews these experiences in order to draw lessons on the benefits of an institutionalized approach to IE, the conditions that favor it, the challenges limiting progress, and some of the important steps in the process of developing such an approach.

Although this chapter focuses on impact evaluation, it is emphasized that IE is

2. See for example: Michael Bamberger and Nobuko Fujita (2008) *Impact Evaluation of Development Assistance*. FASID Evaluation Handbook.

3. Two World Bank publications have discussed the different ways in which impact evaluations have contributed to development management (Bamberger, Mackay and Ooi 2004 and 2005).

only one of many types of evaluation that planners and policymakers use. The institutionalization of IE can only be achieved when it is part of a broader M&E system. It would not make sense, or even be possible, to focus exclusively on IE without building up the monitoring and other data-collection systems on which IE relies. Although IEs are often the most discussed (and most expensive) evaluations, they only provide answers to certain kinds of questions; and for many purposes, other kinds of evaluation will be more appropriate. Consequently, there is a need to institutionalize a comprehensive monitoring and evaluation (M&E) system that provides a menu of evaluations to cover all the information needs of managers, planners, and policymakers.

2. The Importance of Impact Evaluation for ODA Policy and Development Management

The primary goals of ODA programs are to contribute to reducing poverty, promoting economic growth and achieving sustainable development. In order to assess the effectiveness of ODA programs in contributing to these goals it is important to conduct a systematic analysis of development effectiveness. Two common ways to do this are through the development of monitoring and evaluation (M&E) systems or the use of Results-Based Management (Kusek and Rist 2004). While both of these approaches are very valuable, they only measure changes in the conditions of the target population (beneficiaries) that the ODA interventions are intended to affect, and they normally do not include a comparison group (counterfactual) not affected by the program intervention. Consequently it is difficult, if not impossible, to determine the extent to which the observed changes can be attributed to the effects of the project and not to other unrelated factors such as changes in the local or national economy, changes in government policies (such as minimum salaries), or similar programs initiated by the government, other donors or NGOs. An unbiased estimate of the impacts of ODA programs requires the use of a *counterfactual* that can isolate the changes attributable to the program from the effect of these other factors. The purpose of impact evaluation (IE) methodology is to provide rigorous and unbiased estimates of the true impacts of ODA interventions.

Why is the use of rigorous IE methodologies important? Most assessments of ODA effectiveness are based either on data that is only collected from project beneficiaries *after* the project has been implemented, or on the use of M&E or Results-Based Management to measure changes that have

taken place in the target population over the life of the project. In either case data is only generated on the target population and no information is collected on the population that does not benefit from the project, or that in some cases may even be worse off as a result of the project (see Box 1). With all of these approaches there is a strong tendency for the evaluation to have a *positive bias* and to over-estimate the true benefits or effects produced by the project. Typically only project beneficiaries and the government and NGOs actively involved in the project are interviewed, and in most cases they will have a positive opinion of the project (or will not wish to criticize it publicly). None of the families or communities that do not benefit are interviewed and the evaluation does not present any information on the experiences or opinions of these non-beneficiary groups. As a result ODA agencies are mainly receiving positive feedback and they are led to believe that their projects are producing more benefits than is really the case. As the results of most evaluations are positive, the ODA agencies do not have any incentive to question the methodological validity of the evaluations – most of which are methodologically weak and often biased. Consequently there is a serious risk that ODA agencies may continue to fund programs that may be producing much lower impacts than are reported and that may even be producing negative consequences for some sectors of the target population. So in a time of economic crisis when ODA funds are being reduced, and there is a strong demand from policymakers to assess aid effectiveness, there is a real risk that unless impact evaluation methodologies are improved, ODA resources are not being allocated in the most cost-effective way.

Why are so few rigorous impact evaluations commissioned?

Many evaluation specialists estimate that rigorous impact evaluation designs (see following sections) are probably only used in a maximum of 10 percent of ODA impact evaluations, so that for the other 90 percent (or more) there is a serious risk of misleading or biased estimates of the impact and effectiveness of ODA assistance. Given the widespread recognition by ODA agencies of the importance of rigorous impact evaluations, why are so few rigorous impact evaluations conducted? There are many reasons, including the limited evaluation budgets of many agencies, and the fact that most evaluations are not commissioned until late in the project cycle and that consultants are only given a very short time (often less than two weeks) to conduct data collection. Also many government agencies see evaluation as a threat or something that will demand a lot of management time without producing useful find-

ings, and many ODA agencies are more concerned to avoid critical findings that might create tensions with host country agencies or prejudice future funding, than they are to ensure a rigorous and impartial assessment of the programs. Also, as many evaluations produce a *positive bias* (see Box 1) and show programs in a positive light (or under-emphasize negative aspects),

Box 1. The danger of over-estimating project impact when the evaluation does not collect information on comparison groups who have not benefited from the project intervention

For reasons of budget and time constraints, a high proportion of evaluations commissioned to assess the effectiveness and impacts of ODA projects only interview project beneficiaries and the agencies directly involved in the projects. When this happens there is a danger of a positive bias in which the favorable effects of the interventions are over-estimated, and the negative consequences are ignored or under-estimated. However, if these negative effects and impacts are taken into account, the net positive impact of the project on the total intended target population may be significantly reduced.

- An evaluation of a micro-credit program promoting the manufacture and marketing of traditional carpets in Bolivia, interviewed a sample of carpet makers who had received loans. It was found that on average their income from the sale of carpets had increased significantly. It was concluded that microcredit was an effective instrument for increasing the income of traditional artisans and reducing poverty. However, when carpet manufacturers who had not received loans were interviewed it was found that on average their income had declined. One of the reasons was that loan recipients were able to rent or purchase a vehicle to get their carpets to market more quickly and cheaply, which gave them a competitive advantage. On the basis of this additional information it was estimated that the total sales of carpets had probably not increased very much, but that a larger share of the market was now controlled by loan recipients. The inclusion of a control group (artisans who did not receive loans) radically changed the conclusions of the evaluation.
- Consultants commissioned to evaluate the impacts of food-for-work programs on women's economic and social empowerment were only able to spend an average of 3-4 days visiting project locations and meeting with affected communities in each country. Typically the consultants met with the local government agencies managing the projects, the local NGOs responsible for implementing the projects and residents of the communities where the programs operated. In each community consultants met with groups of women participating in the food-for-work programs and with many of their husbands. It was apparent that the project had produced significant increases in the income of the women, that their husbands were very supportive of the economic activities of their wives and that there was convincing evidence of the women's economic empowerment and increase in their self-confidence and social empowerment. In most cases the evaluation ended at this point and a very positive report was produced. However, in several cases the consultants also contacted key informants not involved in the project in order to obtain information on the experiences of other women who had not participated in the project. For example, a local nurse who regularly visited women in these same communities reported that many women who had attended the initial meetings had been forbidden by their husbands to participate in the project and in quite a few cases had been beaten for attending without his permission. Many men were unemployed and felt humiliated that they were not able to fulfill their traditional role of providing economically for their family. This again illustrates that a completely different image of the project was obtained when an effort was made to obtain information on the situation of non-participants and not to simply base the evaluation report on meetings with those who had benefited.

many agencies do not feel the need for more rigorous (as well as more expensive and time-consuming) evaluation methodologies. One of the challenges for the institutionalization of impact evaluation is to convince both ODA agencies and host country governments that rigorous and objective impact evaluations can become valuable budgetary, policy and management tools.

3. Defining Impact Evaluation

The primary purpose of an Impact Evaluation (IE) is to estimate the magnitude and distribution of changes in outcome and impact indicators among different segments of the target population, and the extent to which these changes can be attributed to the interventions being evaluated. In other words, is there convincing evidence that the intervention being evaluated has contributed towards the achievement of its intended objectives? IE can be used to assess the impacts of *projects* (a limited number of clearly defined and time-bound interventions, with a start and end date, and a defined funding source); *programs* (broader interventions that often comprise a number of projects, with a wider range of interventions and a wider geographical coverage and often without an end date); and *policies* (broad strategies designed to strengthen or change how government agencies operate or to introduce major new economic, fiscal, or administrative initiatives). IE methodologies were originally developed to assess the impacts of precisely defined interventions (similar to the project characteristics described above); and an important challenge is how to adapt these methodologies to evaluate the multi-component, multi-donor sector and country-level support packages that are becoming the central focus of development assistance.

A well-designed IE can help managers, planners, and policymakers *avoid* continued investment in programs that are not achieving their objectives, *avoid* eliminating programs that either are or potentially could achieve their objectives, *ensure* that benefits reach all sectors of the target population, *ensure* that programs are implemented in the most efficient and cost-effective manner and that they *maximize* both the quantity and the quality of the services and benefits they provide, and *provide* a decision tool for selecting the best way to invest scarce development resources. Without a good IE, there is an increased risk of reaching wrong decisions on whether programs should continue or be terminated and how resources should be allocated.

Two different definitions of impact evaluation are widely used. The first, that can be called the *technical* or *statistical definition* defines IE as an evaluation that

... assesses changes in the well-being of individuals, households, communities or firms that can be attributed to a particular project, program, or policy. The central impact evaluation question is what would have happened to those receiving the intervention if they had not in fact received the program. Since we cannot observe this group both with and without the intervention, the key challenge is to develop a counterfactual—that is, a group which is as similar as possible (in observable and unobservable dimensions) to those receiving the intervention. This comparison allows for the establishment of definitive causality—attributing observed changes in welfare to the program, while removing confounding factors.

[Source: World Bank PovertyNet website ⁴]

The second, that can be called the *substantive long-term effects* definition is espoused by the Organization for Economic Co-operation and Development's Development Assistance Committee (OECD/DAC). This defines impact as:

positive and negative, primary and secondary, long-term effects produced by a development intervention, directly or indirectly, intended or unintended

[Source: OECD-DAC 2002, p. 24].

While the OECD/DAC definition does not require a particular methodology for conducting an IE, but does specify that impact evaluations only assess *long-term effects*; the *technical* definition requires a particular methodology (the use of a counterfactual, based on a pretest/posttest project/control group comparison) but does not specify a time horizon over which impacts should be measured, and does not specify the kinds of changes (outputs, outcomes or impacts) that can be assessed. To some extent these definitions are inconsistent as the *technical* definition would permit an IE to be conducted at any stage of the project cycle as long as a counterfactual is

4. For extensive coverage of the technical/statistical definition of IE and a review of the main quantitative analytical techniques, see the World Bank's Development Impact Evaluation Initiative Web site: <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20205985~menuPK:435951~pagePK:148956~piPK:216618~theSitePK:384329,00.html>. For an overview of approaches used by IEG, see White (2006), and for a discussion of strategies for conducting IE (mainly at the *project* level) when working under budget, time, and data constraints see Bamberger (2006).

used; while the *substantive* definition would only permit an IE to assess long-term effects but without specifying any particular methodology. This distinction between the two definitions has proved to be important as many evaluators argue that impacts can be estimated using a number of different methodologies (the *substantive* definition), whereas advocates of the *technical* definition argue that impacts can only be assessed using a limited number of statistically strong IE designs and that randomized control trials should be used wherever possible. Box 2 explains that in this paper we will use a comprehensive definition of IE that encompasses both the *technical* and *substantive* definitions.

Box 2. Two alternative definitions of impact evaluation [IE]

There are two widely used definitions of IE. The first *technical* definition, supported by most economists and quantitative researchers, requires that the evaluation design include a counterfactual (a comparison group that closely matches the characteristics of the project or treatment group). This definition can cover both evaluations of short-term outputs or outcomes that are measured while a project is still being implemented as well as evaluations of long-term impacts several years after the project has been completed. The second *Substantive Long-term effects* definition, proposed by OECD-DAC, states that IE can only be used to measure long-term impacts after a project has been operating for some time, but does not specify a particular methodology.

The two definitions are not completely compatible. The *technical* definition specifies the methodology but not the time-horizon, while the *substantive* definition specifies the time-horizon but recognizes the possibility of using different evaluation methodologies. In this chapter we will use a *comprehensive* definition of impact evaluation that encompasses both of these definitions. One of the important decisions that development agencies and developing country governments must make is whether to restrict their impact evaluations to only one of the two definitions or to recognize the benefits of using both definitions for different kinds of evaluations. This is an important decision, because as we shall see, it is only possible to use statistically rigorous quantitative IE designs that satisfy the *technical* definition for a relatively small number of evaluations.

Although IE is the most frequently discussed type of program evaluation, it is only one of many types of evaluation that provide information to policymakers, planners, and managers at different stages of a project or program cycle. Table 1 lists a number of different kinds of program evaluations that can be commissioned during project planning, while a project is being implemented, at the time of project completion, or after the project has been operating for some time. Although many impacts cannot be fully assessed until an intervention has been operating for several years, planners and policymakers cannot wait three or five years before receiving feedback. Consequently, many IEs are combined with formative or process evaluations designed to provide preliminary findings while the project is still being implemented to

Table 1. Types of project or program evaluation used at different stages of the project cycle

1. Project planning	2. Implementation	3. Project completion	4. Post completion
a. Exploratory studies	a. Evaluations based on program theory (logic) models	a. End-of-project survey for preliminary impact evaluation**	a. Post-test data collection for pre-test/post-test impact evaluation**
b. Project appraisal	b. Results-based M&E framework**	b. Implementation completion report	b. Post-test impact assessment**
c. Beneficiary assessment studies	c. Monitoring reports	c. Repeated cost-benefit or cost-effectiveness analysis	c. Sustainability assessment
d. Baseline data collection	d. Beneficiary assessment	d. Results-Based Management Project Assessment	d. Public expenditure tracking study (study repeated to assess improvements)
e. Pre-test data collection for pre-test/post-test impact evaluation**	e. Rapid feedback studies	e. Historical reconstruction	
f. Cost benefit and cost-effectiveness analysis	f. Process analysis**		
	g. Public expenditure tracking studies		
	h. Action research		
	i. Mid-term review		

Note: ** program evaluations that can be part of an impact evaluation design

assess whether a program is on track and likely to achieve its intended outcomes and impacts.

The most widely-used IE designs

There is no one design that fits all IE. The best design will depend on what is being evaluated (a small project, a large program, or a nationwide policy); the purpose of the evaluation; budget, time, and data constraints; and the time horizon (is the evaluation designed to measure medium- and long-term impacts once the project is completed or to make initial estimates of potential future impacts at the time of the midterm review or the implementation completion report?). IE designs can also be classified according to their level of statistical rigor (see Table 2). The most rigorous designs, from the statistical point of view are the experimental designs, commonly known as randomized control trials (Design 1 in Table 2). These are followed, in descending order of statistical rigor by strong quasi-experimental designs that use pre-test/post-test control group designs (Designs 2-4); and weaker quasi-experimental designs where baseline data has not been collected on either or both of the project and control groups (Designs 5-8).

The least statistically rigorous are the non-experimental designs (Designs 9-10) that do not include a control group and that may also not include baseline data on the project group. According to the *technical* definition the non-experimental designs should not be considered as IE because they do not include a counterfactual (control group); but according to the *substantive* def-

inition these can be considered IE when they are used to assess the long-term project outcomes and impacts. However, a critical factor in determining the methodological soundness of non-experimental designs is the adequacy of the alternative approach proposed to examine causality in the absence of a conventional counterfactual ⁵.

Advocates of the *technical* definition of an IE often claim that randomized control trials and strong quasi-experimental designs are the “best” and “strongest” designs (some use the term the “gold standard”). However, it is important to appreciate that these designs should only be considered as the “strongest” in an important but narrow statistical sense as their strength lies in their ability to eliminate or control for selection bias. While this is an extremely important advantage, critics point out that these designs are not necessarily stronger than other designs with respect to other criteria (such as construct validity, the validity and reliability of indicators of outcomes and impacts, and the evaluators’ ability to collect information on sensitive topics and to identify and interview difficult-to-reach groups). When used in isolation these “strong” designs, also have some fundamental weaknesses such as ignoring the process of project implementation and lack of attention to the local context in which each project is implemented.

It is important for policymakers and planners to keep in mind that there are relatively few situations in which the most rigorous evaluation designs (Designs 1-4) can be used ⁶. While there is an extensive evaluation literature on the small number of cases where strong designs have been used, much less guidance is available on how to strengthen the methodological rigor of the majority of IEs that are forced by budget, time, data, or political constraints to use methodologically weaker designs ⁷.

Deciding when an IE is needed and when it can be conducted

IE may be required when policymakers or implementing agencies need to make decisions or obtain information on one or more of the following:

-
5. See Scriven (2009) and Bamberger, Rugh and Mabry (2006) for a discussion of some of the alternative ways to assess causality. Concept Mapping (Kane and Trochim, 2007) is often cited as an alternative approach to the analysis of causality.
 6. Although it is difficult to find statistics, based on discussions with development evaluation experts, this report estimates that randomized control trials have been used in only 1–2 percent of IEs; that strong quasi-experimental designs are used in less than 10 percent, probably not more than 25 percent include baseline surveys, and at least 50 percent and perhaps as high as 75 percent do not use any systematic baseline data.
 7. See Bamberger (2006a), and Bamberger, Rugh & Mabry (2006) for a discussion of strategies for enhancing the quality of impact evaluations conducted under budget, time and data constraints.

- To what extent and under what circumstances could a successful pilot or small-scale program be replicated on a larger scale or with different population groups?
- What has been the contribution of the intervention supported by a single donor or funding agency to a multi-donor or multiagency program?
- Did the program achieve its intended effects, and was it organized in the most cost-effective way?
- What are the potential development contributions of an innovative new program or treatment?

The size and complexity of the program, and the type of information required by policymakers and managers, will determine whether a rigorous and expensive IE is required, or whether it will be sufficient to use a simpler and less expensive evaluation design. There are also many cases where budget and time-constraints make it impossible to use a very rigorous design (such as designs 1-4 in Table 2) and where a simpler and less rigorous design will be the best option available. There is no simple rule for deciding how much an evaluation should cost and when a rigorous and expensive IE may be justified. However, an important factor to consider is whether the benefits of the evaluation (for example, money saved by making a correct decision or avoiding an incorrect one) are likely to exceed the costs of conducting the evaluation. An expensive IE that produces important improvements in program performance can be highly cost-effective; and even minor improvements in a large-scale program may result in significant savings to the government. Of course, it is important to be aware that there are many situations in which an IE is not the right choice and where another evaluation design is more appropriate.

Table 2. Widely used impact evaluation designs ⁸

T = Time	Start of project [pre-test]	Project intervention [a process or a discrete event]	Mid-term evaluation	End of project or when the project has been operating for some time [post-test]	The stage of the project cycle at which each evaluation design begins ⁹
P = Project participants; C = Control Group Ph = Phase P ₁ , P ₂ , C ₁ , C ₂ = First and second observations X = Intervention (An intervention is usually a process, but could be a discrete event.)	T ₁		T ₂	T ₃	
EXPERIMENTAL (RANDOMIZED) DESIGN [Both technical and substantive definitions apply]					
1. Randomized control trials. Subjects are randomly assigned to the project (treatment) and control groups.	P ₁ C ₁	X		P ₂ C ₂	Start
QUASI-EXPERIMENTAL DESIGNS [Both technical and substantive definitions apply]					
Methodologically strong designs with pretest/posttest project and control groups					
2. Pre-test post-test control group design with statistical matching of the two groups. Participants are self-selected or selected by the project agency. Statistical techniques (such as propensity score matching) use secondary data to match the two groups on relevant variables.	P ₁ C ₁	X		P ₂ C ₂	Start
3. Regression discontinuity. A clearly defined cut-off point is used to define project eligibility. Groups above and below the cut-off are compared using regression analysis.	P ₁ C ₁	X		P ₂ C ₂	Start
4. Pre-test post-test control group design with judgmental matching of the two groups. Similar to Design 2 but comparison areas are selected judgmentally with subjects randomly selected from within these areas. Subjects not receiving benefits until Phase 2 can be used as the control for Phase 2 etc..	P ₁ C ₁	X		P ₂ C ₂	Start
Methodologically weaker designs where baseline data has not been collected on the project and/or control group.					
5. Pipeline control group design. When a project is implemented in phases, subjects in Phase 2 (i.e., who will not receive benefits until some later point in time) can be used as the control group for Phase 1 subjects. Subjects not receiving benefits until Phase 3 can be used as the control for Phase 2 etc.	Ph1[P ₁] Ph2[C ₁]	X	Ph1[P ₂] Ph2[P ₂] Ph3[C ₁]	Ph2[P ₂] Ph3[P ₂]	Start of project or when subsequent phase begins
6. Pre-test/post-test control where the baseline study is not conducted until the project has been underway for some time (most commonly this is around the mid-term review)		X	P ₁ C ₁	P ₂ C ₂	During project implementation - often at mid-term
7. Pre-test post-test control of project group combined with post-test comparison of project and control group	P ₁	X		P ₂ C ₂	Start
8. Post-test control of project and control groups		X		P ₁ C ₁	End
NON-EXPERIMENTAL DESIGNS [Only the substantive definition applies]					
9. Pre-test post-test comparison of project group	P ₁	X		P ₂	Start
10. Post-test analysis of project group		X		P ₁	End

8. Each design category indicates whether it corresponds to the *technical* definition, the *substantive* definition or both. The *substantive* definition can be applied to every design category but only in cases where the project has been operating for sufficient time for it to be possible to measure long-term effects.

9. Baseline data may be obtained either from the collection of new data through surveys or other data collection instruments, or from secondary data – census or survey data that has already been collected. When secondary data sources are used to reconstruct the baseline, the evaluation may not actually start until late in the project cycle but the design is classified as a pretest/posttest design.

4. Institutionalizing Impact Evaluation

Institutionalization of IE at the sector or national level occurs when

- The evaluation process is country-led and managed by a central government ministry or a major sectoral agency
- There is strong “buy-in” to the process from key stakeholders
- There are well-defined procedures and methodologies
- IE is integrated into sectoral and national M&E systems that generate much of the data used in the IE studies
- IE is integrated into national budget formulation and development planning
- There is a focus on evaluation capacity development (ECD)

Institutionalization is a process, and at any given point it is likely to have advanced further in some areas or sectors than in others. The way in which IE is institutionalized will also vary from country to country, reflecting different political and administrative systems and traditions and historical factors such as strong donor support for programs and research in particular sectors. It should be pointed out that the institutionalization of IE is a special case of the more general strategies for institutionalizing an M&E system, and many of the same principles apply. As pointed out earlier, IE can only be successfully institutionalized as part of a well-functioning M&E system.

Although the benefits of well-conducted IE for program management, budget management and policymaking are widely acknowledged, the contribution of many IE to these broader financial planning and policy activities has been quite limited because the evaluations were selected and funded in a somewhat ad hoc and opportunistic way that was largely determined by the interests of donor agencies or individual ministries. The value of IEs to policymakers and budget planners can be greatly enhanced once the selection, dissemination and use of the evaluations becomes part of a national or sector IE system. This requires an annual plan for selection of the government’s priority programs on which important decisions have to be made concerning continuation, modification, or termination and where the evaluation framework permits the comparison of alternative interventions in terms of potential cost-effectiveness and contribution to national development goals. The examples presented in the following sections illustrate the important benefits that have been obtained in countries where significant progress has been made toward institutionalization.

Alternative pathways to the institutionalization of IE¹⁰

There is no single strategy that has always proved successful in the institutionalization of IE. Countries that have made progress have built on existing evaluation experience, political and administrative traditions, and the interest and capacity of individual ministries, national evaluation champions, or donor agencies. Although some countries—particularly Chile—have pursued a national M&E strategy that has evolved over a period of more than 30 years; most countries have responded in an ad hoc manner as opportunities have presented themselves.

Figure 1 identifies three alternative pathways to the institutionalization of IE that can be observed. **The first pathway** (the ad hoc or opportunistic approach) evolves from individual evaluations that were commissioned to take advantage of available funds or from the interest of a government official or a particular donor. Often evaluations were undertaken in different sectors, and the approaches were gradually systematized as experience was gained in selection criteria, effective methodologies, and how to achieve both quality and utilization. A central government agency—usually finance or planning—is either involved from the beginning or becomes involved as the focus moves toward a national system. Colombia's national M&E system, SINERGIA, is an example of this pathway (Box 3.)

Box 3. Colombia: Moving from the Ad Hoc Commissioning of IE by the Ministry of Planning and Sector Ministries toward Integrating IE into the National M&E System (SINERGIA)

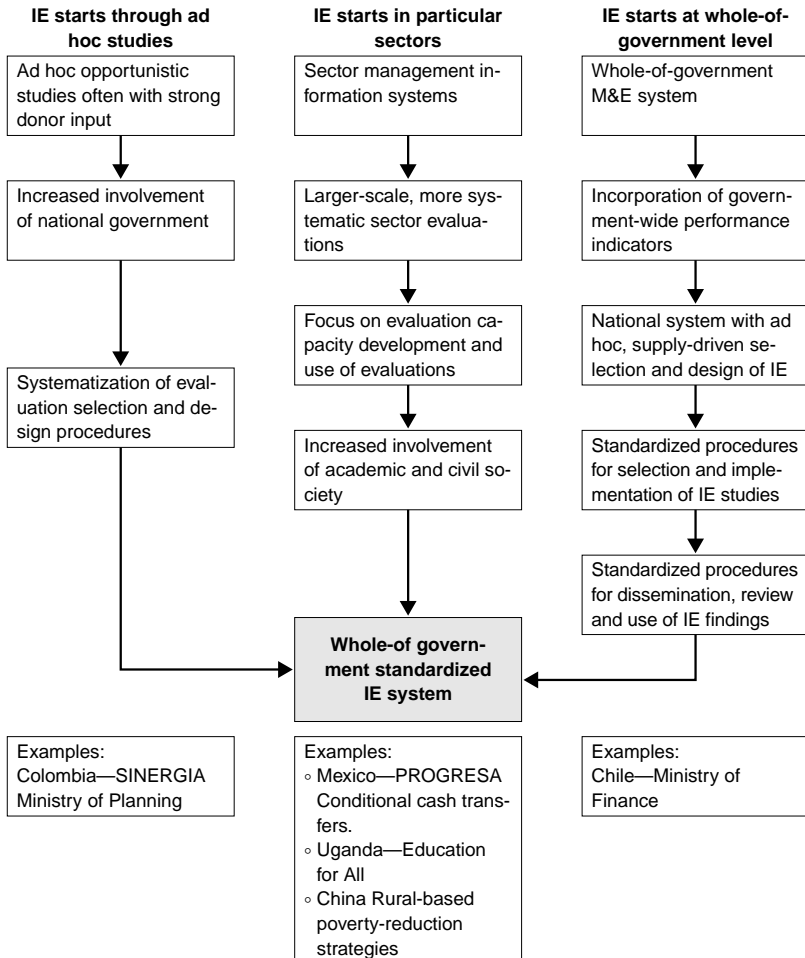
In Colombia the Ministry of Planning is responsible for managing the National System for Evaluation of Public Sector Performance (SINERGIA). The most visible and heavily utilized component is the subsystem for monitoring progress against a total of 320 country development and presidential goals.

Although IE was initiated in 1999, these goals have evolved since 2000 to be commissioned and managed from SINERGIA for a wide range of priority government programs. To date, SINERGIA has played a major role in the selection of the programs to be evaluated. Initially it was a somewhat ad hoc process—partly determined by the interest of international funding agencies. As the program of IE evolved, the range of methodologies was broadened and technical criteria in the selection of programs to be evaluated were formalized through policy documents (with more demand-side involvement from the agencies managing the programs being evaluated) and in how the findings are used. Most of the IEs carried out use rigorous econometric evaluation techniques. A World Bank loan is supporting the strengthening of the system with specific activities aiming to further institutionalize IE.

Source: Mackay (2007, pp. 31–36), Bamberger 2009.

The second pathway is where IE expertise is developed in a priority sector supported by a dynamic government agency and with one or more

Figure 1. Three Pathways for the Evolution of Institutionalized IE Systems*



* *Note:* The institutionalized systems may employ either or both the technical or the substantive definitions of IE. The systems also vary in terms of the level of methodological rigor [which of the designs in table 2] that they use.

champions, and where there are important policy questions to be addressed and strong donor support. Once the operational and policy value of these

- Most of the examples of institutionalization of IE in this paper are taken from Latin American because this is considered to be the region where most progress has been made. Consequently most of the literature on institutionalization of IE mainly cites examples from countries such as Mexico, Colombia and Chile. Progress has undoubtedly been made in many Asian countries as well as other countries, but further research is required to document these experiences.

evaluations has been demonstrated, the sectoral experience becomes a catalyst for developing a national system. The evaluations of the PROGRESA conditional cash transfer programs in Mexico are an example of this approach (Box 4).

Box 4. Mexico: Moving from an Evaluation System Developed in One Sector toward a National Evaluation System (SEDESOL)

In Mexico a series of rigorous evaluations of the Progresas Conditional Cash Transfer programs were conducted over a number of years. The evaluations convincingly demonstrated the effectiveness of conditional cash transfers as a way to improve the welfare (particularly education and health) of large numbers of low-income families. The evaluations are considered to have been a major contributing factor in convincing the new government that came to power in 2002 to continue these programs, which had been started by the previous administration. The evaluations also served to convince policy makers of the technical feasibility and policy value of rigorous IEs and contributed to the passing of a law by Congress in 2007 mandating the evaluation of all social programs. This law also created the National Commission for the Evaluation of Social Programs, which was assigned the responsibility for regulating the development of monitoring and evaluation functions in the social sectors. A similar continuity was achieved in Colombia, where progress is also being made toward a national M&E system (see Box 3).

Source: Mackay (2007, p. 56), Bamberger 2009.

Experience suggests that IE can evolve at the ministerial or sector level in one of the following ways. Sometimes IE begins as a component built into an existing ministry or sector-wide M&E system. In other cases it is part of a new M&E system being developed under a project or program loan funded by one or several donor agencies. It appears that many of these IE initiatives have failed because they tried to build a stand-alone M&E system into an individual project or program when no such system existed in other parts of the ministry or executing agency. This has not proved an effective way to design an IE system, both because some of the data required for the IE must be generated by the M&E system that is still in process of development, and also because the system is “time bound,” with funding ending at the closing of the project loan—which is much too early to assess impacts.

In many other cases, individual IEs are developed as stand-alone initiatives where either no M&E system exists or if such a system does exist, it is not utilized by the IE team, which generates its own databases. Stand-alone IE can be classified into evaluations that start at the beginning of the project and collect baseline data on the project and possibly a comparison group; evaluations that start when the project is already under way—possibly even nearing completion; and those that are not commissioned until the project has ended.

The evaluations of the national *Education for All* program in Uganda offer a second example of the sector pathway ¹¹. These evaluations increased interest in the existing national M&E system (the National Integrated M&E System, or NIMES) and encouraged various agencies to upgrade the quality of the information they submit. The World Bank Africa Impact Evaluation Initiative (AIM) is an example of a much broader regional initiative—designed to help governments strengthen their overall M&E capability and systems through sectoral pathways—that is currently supporting some 90 experimental and quasi-experimental IEs in 20 African countries in the areas of education, HIV, malaria, and community-driven development (see Box 5). Similarly, at least 40 countries in Asia, Latin America, and the Middle East are taking sectoral approaches to IE with World Bank support. A number of similar initiatives are being promoted through recently created international collaborative organizations such as the Network of Networks for Impact Evaluation (NONIE) ¹² and the International Initiative for Impact Evaluation (3IE) ¹³.

According to Ravallion (2008), China provides a dramatic example of the large-scale and systematic institutionalization over more than a decade of IE as a policy instrument for testing and evaluating potential rural-based poverty-reduction strategies. In 1978 the Communist Party's 11th Congress adopted a more pragmatic approach whereby public action was based on demonstrable success in actual policy experiments on the ground:

A newly created research group did field work studying local experiments on the de-collectivization of farming using contracts with individual farmers. This helped convince skeptical policy makers ... of the merits of scaling up the local initiatives. The rural reforms that were then implemented nationally helped achieve probably the most dramatic reduction in the extent of poverty the world has yet seen (Ravallion 2008, p. 2, ... indicates text shortened by the present author).

11. A video of a presentation on this evaluation made during the World Bank January 2008 Conference on "Making smart policy: using impact evaluation for policymaking" is available at: <http://info.worldbank.org/etools/BSPAN/PresentationView.asp?PID=2257&EID=1006>

12. <http://www.worldbank.org/ieg/nonie/index.html>

13. <http://www.3ieimpact.org>

Box 5. Africa Impact Evaluation Initiative

The Africa Impact Evaluation Initiative (AIM) is a program of the World Bank's Africa Region. AIM uses a sector approach to generating and supporting impact evaluations. AIM currently houses umbrella initiatives in the following thematic areas: Education, HIV, Malaria, and Community-Driven Development. Each thematic area is coordinated by a team that provides organizational and technical advisory services to the participating country IE teams.

How Does AIM differ from other Impact Evaluation Initiatives?

Builds Capacity of Governments to Implement Impact Evaluation

- *Complements the government's efforts to think critically about education policies* and identifies policy questions of importance through in-country clinics and cross-country workshops. The cross-country venues allow for dissemination of information, resource sharing, and peer-to-peer learning regarding the efficacy of education interventions
- *Generates options for alternative interventions* to address policy questions through distilling lessons from successful IEs in selected thematic areas
- *Builds the government's capacity to rigorously test the impact of interventions* with experimental or quasi-experimental methods using a learning-by-doing approach. The process includes working with the government team to:
 - Create learning teams within the relevant Ministry
 - Articulate the design, implementation plan, timeline and budget of IE
 - Collect data, process data, summarize findings and disseminate results

Changes the Way Decisions are Made: Supporting Governments to Manage for Results

AIM improves governments' ability to manage for results by providing rigorous evidence on the effectiveness of interventions and sharpening the project implementation process.

This is a natural result of country-led prospective IEs necessitating that development of the following during project implementation:

- Clearly articulated policy objective
- Articulated link between project design and desired outcome
- Rigorously adhered-to implementation and data collection plan

Ensures Demand-Driven Knowledge by Using Bottom-Up Approach to Learning What Works

- Identifies successful interventions collecting results from country-level, country-driven impact evaluations across a diverse set of African countries. Each thematic area has a technical advisory group providing a practicable framework to harmonize approaches to key areas within IE.
- Enriches the evidence to inform policy decisions and program design throughout the region by disseminating information in an easily understood format through the AIM website, seminars, workshops, and government presentations.

For more information: <http://worldbank.org/afr/impact>.

The third pathway is where a planned and integrated series of IEs was developed from the start as one component of a whole-of-government system, managed and championed by a strong central government agency, usually the ministry of finance or planning. Chile is a good example of a national M&E system in which there are clearly defined criteria and guidelines for

the selection of programs to be evaluated, their conduct and methodology, and how the findings will be used (Box 6).

Box 6. Chile: Rigorous IEs Introduced as Part of an integrated Whole-of-Government M&E System

The government of Chile has developed over the past 14 years a whole-of-government M&E system with the objective of improving the quality of public spending. Starting in 1994, a system of performance indicators was developed; rapid evaluations of government programs were incorporated in 1996; and in 2001 a program of rigorous impact evaluations was incorporated. There are two clearly defined IE products. The first are rapid ex post evaluations that follow a clearly defined and rapid commissioning process, where the evaluation has to be completed in less than 6 months for consideration by the Ministry of Finance as part of the annual budget process. The second are more comprehensive evaluations that can take up to 18 months and cost \$88,000 on average.

The strength of the system is that it has clearly defined and cost-effective procedures for commissioning, conducting, and reporting of IEs, a clearly defined audience (the Ministry of Finance), and a clearly understood use (the preparation of the annual budget). The disadvantages are that the focus of the studies is quite narrow (only covering issues of interest to the Ministry of Finance) and the involvement and buy-in from the agencies being implemented is typically low. Some have also suggested that there may be a need to incorporate some broader and methodologically more rigorous IEs of priority government programs (similar to the PROGRESA evaluations in Mexico).

Source: Mackay (2007, pp. 25–30). Bamberger 2009

Guidelines for institutionalizing IEs at the national or sector levels

As discussed earlier, IEs often begin in a somewhat ad hoc and opportunistic way, taking advantage of the interest of key stakeholders and available funding opportunities. The challenge is to build on these experiences to develop capacity to select, conduct, disseminate, and use evaluations. Learning mechanisms, such as debriefings and workshops, can also be a useful way to streamline and standardize procedures at each stage of the IE process. It is helpful to develop an IE handbook for agency staff summarizing the procedures, identifying the key decision points, and presenting methodological options (DFID 2005). Table 3 lists important steps in institutionalizing an IE system.

Table 3. Key steps for institutionalizing impact evaluation at the national and sector levels

1. Conduct an initial diagnostic study to understand the context in which the evaluations will be conducted ¹⁴ .
2. The diagnostic study should take account of local capacity, and where this is lacking, it should define what capacities are required and how they can be developed
3. A key consideration is whether a particular IE will be a single evaluation that probably will not be repeated or whether there will be a continuing demand for such IEs.
4. Define the appropriate option for planning, conducting and/or managing the IE, such as: <ul style="list-style-type: none"> • <i>Option 1:</i> Most IE will be conducted by the government ministry or agency itself. • <i>Option 2:</i> IE will be planned, conducted, and/or managed by a central government agency, with the ministry or sector agency only being consulted when additional technical support is required. • <i>Option 3:</i> IE will be managed by the sector agency but subcontracted to local or international consultants. • <i>Option 4:</i> The primary responsibility will rest with the donor agencies.
5. Define a set of standard and transparent criteria for the selection of the IE to be commissioned each year.
6. Define guidelines for the cost of an IE and how many IEs should be funded each year.
7. Clarify who will define and manage the IE.
8. Define where responsibility for IE is located within the organization and ensure that this unit has the necessary authority, resources, and capacity to manage the IEs.
9. Conduct a stakeholder analysis to identify key stakeholders and to understand their interest in the evaluation and how they might become involved ¹⁵ .
10. A steering committee may be required to ensure that all stakeholders are consulted. It is important, however, to define whether the committee only has an advisory function or is also required to approve the selection of evaluations ¹⁶ .
11. Ensure that users continue to be closely involved throughout the process.
12. Develop strategies to ensure effective dissemination and use of the evaluations.
13. Develop an IE handbook to guide staff through all stages of the process of an IE: identifying the program or project to be evaluated, commissioning, contracting, designing, implementing, disseminating, and using the IE findings.
14. Develop a list of prequalified consulting firms and consultants eligible to bid on requests for proposals.

Source: Bamberger 2009. Institutionalizing Impact Evaluation within the Framework of a Monitoring and Evaluation System. Independent Evaluation Group. The World Bank

Integrating IE into sector and/or national M&E and other data-collection systems

The successful institutionalization of IE will largely depend on how well the

14. For a comprehensive discussion of diagnostic studies see Mackay (2007 Chapter 12).

15. Patton (2008) provides guidelines for promoting stakeholder participation.

16. Requiring steering committees to approve evaluation proposals or reports can cause significant delays as well as sometimes force political compromises in the design of the evaluation. Consequently, the advantages of broadening ownership of the evaluation process must be balanced against efficiency.

selection, implementation, and use of IE are integrated into sector and national M&E systems and national data-collection programs. This is critical for several reasons.

First, much of the data required for an IE can be obtained most efficiently and economically from the program M&E systems. This includes information on:

- How program beneficiaries (households, communities, and so forth) were selected and how these criteria may have changed over time
- How the program is being implemented (including which sectors of the target population do or do not have access to the services and benefits), how closely this conforms to the implementation plan, and whether all beneficiaries receive the same package of services and of the same quality
- The proportion of people who drop out, the reasons for this, and how their characteristics compare with people who remained in the program
- How program outputs compare with the original plan

Second, IE findings that are widely disseminated and used provide an incentive for agencies to improve the quality of M&E data they collect and report, thus creating a virtuous circle. One of the factors that often affects the quality and completeness of M&E data is that overworked staff may not believe that the M&E data they collect are ever used, so there is a temptation to devote less care to the reliability of the data. For example, Ministry of Education staff in Uganda reported that the wide dissemination of the evaluations of the Education for All program made them aware of the importance of carefully collected monitoring data, and it was reported that the quality of monitoring reporting improved significantly.

Third, access to monitoring data makes it possible for the IE team to provide periodic feedback to managers and policy makers on interim findings that could not be generated directly from the IE database. This increases the practical and more immediate utility of the IE study and overcomes one of the major criticisms that clients express about IE—namely that there is a delay of several years before any results are available.

Fourth, national household survey programs such as household income and expenditure surveys, demographic and health surveys, education surveys, and agricultural surveys provide very valuable sources of secondary data for strengthening methodological rigor of IE design and analysis (for example, the use of propensity score matching to reduce sample selection

bias). Some of the more rigorous IEs have used cooperative arrangements with national statistical offices to piggy-back information required for the IE onto an ongoing household survey or to use the survey sample frame to create a comparison group that closely matches the characteristics of the project population. Piggy-backing can also include adding a special module. Although piggy-backing can save money, experience shows that the required coordination can make this much more time consuming than arranging a stand-alone data-collection exercise.

5. Creating Demand for IE¹⁷

Efforts to strengthen the governance of IE and other kinds of M&E systems are often viewed as technical fixes—mainly involving better data systems and the conduct of good quality evaluations (Mackay 2007). Although the creation of evaluation capacity needed to provide high-quality evaluation services and reports is important, these *supply-side interventions* will have little effect unless there is sufficient *demand* for quality IE.

Demand for IE requires that quality IEs are seen as an important policy and management tool in one or more of the following areas: (a) assisting resource-allocation decisions in the budget and planning process; (b) to help ministries in their policy formulation and analytical work; (c) to aid ongoing management and delivery of government services; and (d) to underpin accountability relationships.

Creating demand requires that there be sufficiently powerful incentives within a government to conduct IE, to create a good level of quality, and to use IE information intensively. A key factor is to have a public sector environment supportive of the use of evaluation findings as a policy and management tool. If the environment is not supportive or is even hostile to evaluations, raising awareness of the benefits of IE and the availability of evaluation expertise might not be sufficient to encourage managers to use these resources. Table 4 suggests some possible carrots (positive incentives), sticks (sanctions and threats), and sermons (positive messages from key figures) that can be used to promote the demand for IE.

The incentives are often more difficult to apply to IE than to promoting general M&E systems for several reasons. First, IEs are only conducted on

17. This section adapts the discussion by Mackay (2007) on how to create broad demand for M&E to the specific consideration here of how to create demand for IE.

selected programs and at specific points in time; consequently, incentives must be designed to encourage use of IE in appropriate circumstances but not to encourage its overuse—for example, where an IE would be premature or where similar programs have already been subject to an IE. Second, as flexibility is required in the choice of IE designs, it is not meaningful to propose standard guidelines and approaches, as can often be done for M&E. Finally, many IEs are contracted to consultants so agency staff involvement (and consequently their buy-in) is often more limited.

It is important to actively involve some major national universities and research institutions. In addition to tapping this source of national evaluation expertise, universities—through teaching, conferences, research, and consulting—can play a crucial role in raising awareness of the value and multiple uses of IE. Part of the broad-based support for the PROGRESA programs and their research agenda was because they made their data and analysis available to national and international researchers on the Internet. This created a demand for further research and refined and legitimized the sophisticated methodologies used in the PROGRESA evaluations. Both Mexico's PROGRESA and Colombia's *Familias en Accion* recognized the importance of dissemination (through high-profile conferences, publications, and working with the mass media) in demonstrating the value of evaluation and creating future demand.

6. Capacity Development for IE

In recent years there has been a significant awareness on the part of ODA agencies of the importance of evaluation capacity development (ECD). For example, the World Bank's Independent Evaluation Group has an ECD website listing World Bank publications and studies on ECD; OECD-DAC has organized a network providing resources and links on ECD and related resources¹⁸, and many other development agencies also offer similar resources. A recent series of publications by UNICEF presents the broader context within which evaluation capacity must be developed at the country and regional levels; and documents recent progress – including in the development of national and regional evaluation organizations¹⁹.

The increased attention to ECD is largely due to the recognition that past

18. http://www.capacity.org/en/resource_corners/learning/useful_links/oecd_dac_network_on_development_evaluation

Table 4. Incentives for IE—Some Rewards [“Carrots”], Sanctions [“Sticks”], and Positive Messages from Important People [“Sermons”]

Carrots	Sticks	Sermons
<ul style="list-style-type: none"> • Work to build ownership by clients and other stakeholders at all stages of the IE. • Award prizes—high-level recognition of good or best practice IE. • Take a collegiate approach to IE among key ministries. • Provide financial and other incentives to implementing agencies to design administrative reporting systems so that they can easily be used as baseline and process monitoring data for subsequent IEs. • Provide additional funding to ministries to conduct IE. • Provide technical assistance and funding support (via loans) from donors to governments. • Present IE findings in an easily understandable format. • Assist agencies and consultants in conducting IE (help desks, manuals, and other resources). • Ensure that data providers understand how the data will be used and the importance of providing accurate data to enable IEs to be conducted. • Provide IE training programs for managers and staff. • Identify and highlight good practice examples of IE planning and implementation. • Establish a government-wide network of IE staff. • Provide financial support and technical assistance for government IE from multilateral and bilateral donors. 	<ul style="list-style-type: none"> • Enact laws, decrees, or regulations mandating the planning, conduct, and reporting of IE. • Penalize noncompliance with agreed IE recommendations. • Withhold part of funding from ministries/agencies that fail to conduct IE. • Highlight adverse IE information in reports to Parliament/Congress and disseminate widely. • Highlight poor-quality IE planning, data systems, performance indicators, IE techniques, and IE reporting. A supreme audit institution—a central ministry such as finance or the president's office, and possibly internal audit—can play this role. • Highlight IE findings concerning government performance to civil society. 	<ul style="list-style-type: none"> • Solicit high-level statements of endorsement of IE by the president, ministers, heads of ministries, deputies, and so forth. • Hold awareness-raising seminars and workshops to demystify IE, provide comfort about its “do-ability,” and explain what’s in it for participants. • Use actual examples of influential IEs to demonstrate their utility and cost-effectiveness. • Explain to service managers and staff how IE can help them deliver better services to clients. • Pilot some IEs to demonstrate their usefulness. • Hold conferences/seminars on good practice IE systems in particular ministries, in other countries, and so forth. • Establishing a network of officials working on IE—helps showcase good practice examples in ministries, demonstrates their feasibility, and helps encourage quality standards.

Note: The terms in inverted commas in the title were used by Mackay

Source: Adapted from Mackay (2007, Table 11.1).

19. Segone and Ocampo (editors) 2006. *Creating and Developing Evaluation Organizations: Lessons from Africa, Americas, Asia, Australasia and Europe*. UNICEF; Segone (editor) 2008. *Bridging the gap: the role of monitoring and evaluation in evidence-based policymaking*. UNICEF; Segone (editor) 2009. *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results*. UNICEF.

efforts to strengthen, for example, statistical capacity and national statistical data bases were over-ambitious and had disappointing results, often because they focused too narrowly on technical issues (such as statistical seminars for national statistical institutes) without understanding the institutional and other resource constraints faced by many countries. Drawing on this earlier experience, ODA agencies have recognized that the successful institutionalization of IE will require an ECD plan to strengthen the capacity of key stakeholders to fund, commission, design, conduct, disseminate and use IE. On the supply side this involves:

- Strengthening the *supply* of resource persons and agencies able to deliver high-quality and operationally relevant IEs
- Developing the infrastructure for generating secondary data that complement or replace expensive primary data collection. This requires the periodic generation of census, survey, and program-monitoring data that can be used for constructing baseline data and information on the processes of program implementation.

Some of the skills and knowledge can be imparted during formal training programs, but many others must be developed over time through gradual changes in the way government programs and policies are formulated, implemented, assessed, and modified. Many of the most important changes will only occur when managers and staff at all levels gradually come to learn that IE can be helpful rather than threatening, that it can improve the quality of programs and projects, and that it can be introduced without introducing an excessive burden of work.

An effective capacity-building strategy must target at least five main stakeholder groups: agencies that commission, fund, and disseminate IEs; evaluation practitioners who design, implement, and analyze IEs; evaluation users; groups affected by the programs being evaluated; and public opinion. Users include government ministries and agencies that use evaluation results to help formulate policies, allocate resources, and design and implement programs and projects.

Each of the five stakeholder groups requires different sets of skills or knowledge to ensure that their interests and needs are addressed and that IEs are adequately designed, implemented, and used. Some of the broad categories of skills and knowledge described in Table 5 include understanding the purpose of IEs and how they are used; how to commission, finance, and manage IEs; how to design and implement IEs; and the dissemination and use of IEs.

The active involvement of leading national universities and research institutions is also critical for capacity development. These institutions can mobilize the leading national researchers (and also have their own networks of international consultants), and they have the resources and incentives to work on refining existing and developing new research methodologies. Through their teaching, publications, conferences, and consulting, they can also strengthen the capacity of policy makers to identify the need for evaluation and to commission, disseminate, and use findings. Universities, NGOs, and other civil society organizations can also become involved in action research.

An important but often overlooked role of ECD is to help ministries and other program and policy executing agencies design “evaluation-ready” programs and policies. Many programs generate monitoring and other forms of administrative data that could be used to complement the collection of survey data, or to provide proxy baseline data in the many cases where an evaluation started too late to have been able to conduct baseline studies. Often, however, the data are not collected or archived in a way that makes it easy to use for evaluation purposes—often because of simple things such as the lack of an identification number on each participant’s records. Closer cooperation between the program staff and the evaluators can often greatly enhance the utility of project data for the evaluation.

In other cases, slight changes in how a project is designed or implemented could strengthen the evaluation design. For example, there are many cases where a randomized control trial design could have been used, but the evaluators were not involved until it was too late.

There are a number of different formal and less-structured ways evaluation capacity can be developed, and an effective ECD program will normally involve a combination of several approaches. These include formal university or training institute programs ranging from one or more academic semesters to seminars lasting several days or weeks; workshops lasting from a half day to one week; distance learning and online programs; mentoring; on-the-job training, where evaluation skills are learned as part of a package of work skills; and as part of a community development or community empowerment program.

Identifying resources for IE capacity development

Technical assistance for IE capacity development may be available from donor agencies, either as part of a program loan or through special technical

Table 5. IE Skills and Understanding Required by Different Stakeholder Groups

Group	Examples	Evaluation skills and understanding
Funding agencies	<ul style="list-style-type: none"> • Finance ministries and departments • Donor agencies • Foundations • International and large national NGOs 	<ul style="list-style-type: none"> • Defining when IEs are required • Assessing evaluation consultants • Assessing proposals • Estimating IE resource requirements (funds, time, professional expertise) • Preparing terms of reference
Evaluation practitioners	<ul style="list-style-type: none"> • Evaluation units of line ministries • Evaluation departments of ministries of finance and planning • Evaluation units of NGOs • Evaluation consultants • Universities 	<ul style="list-style-type: none"> • Defining client information needs • Adapting theoretically sound designs to real-world budget, time, data, and political constraints • Understanding and selecting from among different evaluation designs • Developing mixed-method approaches • Designing a program theory model • Collecting and analyzing data • Sampling and survey design • Supervising
Evaluation users	<ul style="list-style-type: none"> • Central government agencies (finance, planning, and so on) • Line ministries and executing agencies • NGOs • Foundations • Donor agencies 	<ul style="list-style-type: none"> • Assessing the validity of quantitative evaluation designs and findings • Assessing the adequacy and validity of qualitative and mixed-method designs
Affected populations	<ul style="list-style-type: none"> • Community organizations • Farmers organizations • Trade associations and business groups • Trade unions and workers organizations 	<ul style="list-style-type: none"> • Helping define when evaluations are required • Negotiating with evaluators and funding agencies on the content, purpose, use, and dissemination of evaluations—ensuring that beneficiaries have a voice • Knowing what questions to ask the researchers to include in the evaluation studies • Understanding and using evaluation findings • Conducting participatory evaluations
Public opinion	<ul style="list-style-type: none"> • The general public • The academic community • Civil society 	<ul style="list-style-type: none"> • Knowing how to get evaluations done • Conducting participatory evaluations • Making sure the right questions are asked • Understanding and using evaluation findings

Source: Adapted from Bamberger (2006b, Table 1).

assistance programs and grants. The national and regional offices of United Nations agencies, development banks, bilateral agencies, and NGOs can also provide direct technical assistance or possibly invite national evaluation staff to participate in country or regional workshops. Developing networks with other evaluators in the region can also provide a valuable resource for exchange or experiences or technical advice. Videoconferencing now pro-

vides an efficient and cost-effective way to develop these linkages.

There are now large numbers of Web sites providing information on evaluation resources. The American Evaluation Association, for example, provides extensive linkages to national and regional evaluation associations, all of which provide their own Web sites. The Web site for the World Bank Thematic Group on Poverty Analysis, Monitoring and Impact Evaluation²⁰ provides extensive resources on IE design and analysis methods and documentation on more than 100 IEs. The IEG Web site²¹ also provides extensive resource material on M&E (including IE) as well as links to IEG project and program evaluations.

7. Data Collection and Analysis for IE²²

Data for an IE can be collected in one of the four ways (White 2006):

- From a survey designed and conducted for the evaluation
- By piggy-backing an evaluation module onto an ongoing survey
- Through a synchronized survey in which the program population is interviewed using a specially designed survey, but information on a comparison group is obtained from another survey designed for a different purpose (for example, national household survey)
- The evaluation is based exclusively on secondary data collected for a different purpose, but that includes information on the program and/or potential comparison groups.

The evaluation team should always check for the existence of secondary data or the possibility of coordination with another planned survey (piggy-backing) before deciding to plan a new (and usually expensive) survey. However, it is very important to stress the great benefits from pretest/posttest comparisons of project and control groups in which new baseline data, specifically designed for the purposes of this evaluation, are generated. Though all the other options can produce methodologically sound and operationally useful results and are often the only available option when operating under budget, time, and data constraints, the findings are rarely as strong or useful as when customized data can be produced. Consequently,

20. <http://web.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTDEVIMPEVAINI/0,,menuPK:3998281~pagePK:64168427~piPK:64168435~theSitePK:3998212,00.html>

21. www.worldbank.org/ieg/ecd

22. For a comprehensive review of data collection methods for IE, see <http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTPOVERTY/EXTISPMA/0,,contentMDK:20205985~menuPK:435951~pagePK:148956~piPK:216618~theSitePK:384329,00.html>

the other options should be looked on as a second best rather than as equally sound alternative designs. One of the purposes of institutionalization of IE is to ensure that baseline data can be collected.

Organizing administrative and monitoring records in a way that will be useful for the future IE

Most programs and projects generate monitoring and other kinds of administrative data that could provide valuable information for an IE. However, there is often little coordination between program management and the evaluation team (who are often not appointed until the program has been under way for some time), so much of this information is either not collected or not organized in a way that is useful for the evaluation. When evaluation information needs are taken into consideration during program design, the following are some of the potentially useful kinds of evaluation information that could be collected through the program at almost no cost:

- Program planning and feasibility studies can often provide baseline data on both the program participants and potential comparison groups.
- The application forms of families or communities applying to participate in education, housing, microcredit, or infrastructure programs can provide baseline data on the program population and (if records are retained on unsuccessful applicants) a comparison group.
- Program monitoring data can provide information on the implementation process and in some cases on the selection criteria ²³.

It is always important for the evaluation team to coordinate with program management to ensure that the information is collected and archived in a way that will be accessible to the evaluation team at some future point in time. It may be necessary to request that small amounts of additional information be collected from participants (for example conditions in the communities where participants previously lived or experience with microenterprises) so that previous conditions can be compared with subsequent conditions.

Reconstructing baseline data

The ideal situation for an IE is for the evaluation to be commissioned at the start of the project or program and for baseline data to be collected on the

23. For example, in a Vietnam rural roads project, monitoring data and project administrative records were used to understand the criteria used by local authorities for selecting the districts where the rural roads would be constructed.

project population and a comparison group before the treatment (such as conditional cash transfers, introduction of new teaching methods, authorization of micro-credits, and so forth) begins. Unfortunately, for a number of reasons, an IE is frequently not commissioned until the program has been operating for some time or has even ended. When a posttest evaluation design is used, it is often possible to strengthen the design and analysis by obtaining estimates of the situation before the project began. Techniques for “reconstructing” baseline data are discussed in Bamberger (2006a and 2009) ²⁴.

Using mixed-method approaches to strengthen quantitative IE designs

Most IEs are based on the use of quantitative methods for data collection and analysis. These designs are based on the collection of information that can be counted and ordered numerically. The most common types of information are structured surveys (household characteristics, farm production, transport patterns, access to and use of public services, and so forth); structured observation (for example, traffic counts, people attending meetings, and the patterns of interaction among participants); anthropometric measures of health and illness (intestinal infections and so on); and aptitude and behavioral tests (literacy and numeracy, physical dexterity, visual perception). Quantitative methods have a number of important strengths, including the ability to generalize from a sample to a wider population and the use of multivariate analysis to estimate the statistical significance of differences between the project and comparison group. These approaches also strengthen quality control through uniform sample selection and data-collection procedures and by extensive documentation on how the study was conducted.

However, from another perspective, these strengths are also weaknesses, because the structured and controlled method of asking questions and recording information ignores the richness and complexity of the issues being studied, the context in which data are collected or in which the programs or phenomena being studied operate. An approach that is rapidly gaining in popularity is mixed-method research that seeks to combine the strengths of both quantitative and qualitative designs. Mixed methods recognize that an evaluation requires both *depth of understanding* of the subjects

24. Techniques include using monitoring data and other project documentation and identifying and using secondary data, recall, and participatory group techniques such as focus groups and participatory appraisal.

and the programs and processes being evaluated and *breadth of analysis* so that the findings and conclusions can be quantified and generalized. Mixed-method designs can potentially strengthen the validity of data collection and broaden the interpretation and understanding of the phenomena being studied. It is strongly recommended that all IE should consider using mixed-method designs, as all evaluations require an understanding of both qualitative and quantitative dimensions of the program (Bamberger, Rugh and Mabry, 2006 Chapter 13).

8. Promoting the Utilization of IE

Despite the significant resources devoted to program evaluation, there is widespread concern that—even for evaluations that are methodologically sound—the utilization of evaluation findings is disappointingly limited (Bamberger, Mackay, and Ooi 2004). The barriers to evaluation utilization also affect institutionalization, and overcoming the former will contribute to the latter. There are a number of reasons why evaluation findings are under-utilized and why the process of IE is not institutionalized:

- Lack of ownership
- Lack of understanding of the purpose and benefits of IE
- Bad timing
- Lack of flexibility and responsiveness to the information needs of stakeholders
- Wrong question and irrelevant findings
- Weak methodology
- Cost and number of demands on program staff
- Lack of local expertise to conduct, review, and use evaluations
- Communication problems
- Factors external to the evaluation
- Lack of a supportive organizational environment.

There are additional problems in promoting the use of IE. IE will often not produce results for several years, making it difficult to maintain the interest of politicians and policy makers, who operate with much shorter time-horizons. There is also a danger that key decisions on future program and policy directions will already have been made before the evaluation results are available. In addition, many IE designs are quite technical and difficult to understand.

The different kinds of influence and effects that an IE can have

When assessing evaluation use, it is important to define clearly what is being assessed and measured. For example, are we assessing *evaluation use*—how evaluation findings and recommendations are used by policy makers, managers, and others; *evaluation influence*—how the evaluation has influenced decisions and actions; or *the consequences of the evaluation*. Program or policy outcomes and impacts can also be assessed at different levels: the individual level (for example, changes in knowledge, attitudes or behavior); the implementation level; the level of changes in organizational behavior; and the national-, sector-, or program-level changes in policies and planning procedures.

Program evaluations can be influential in many different ways, not all of which are intended by the evaluator or the client. Table 6 illustrates different kinds of influence that IEs can have.

Ways to strengthen evaluation utilization

Understanding the political context. It is important for the evaluator to understand as fully as possible the political context of the evaluation. Who are the

Table 6. Examples of the Kinds of Influence IEs Can Have

Type of influence	Examples
Providing strong empirical evidence to convince new administrations to continue major programs that were initiated by the previous administration	The <i>Familias en Acción</i> evaluation in Colombia and the PROGRESA evaluation in Mexico**
Helping national budget agencies assess future funding requirements for major government programs	Chile
Alerting public service agencies to problems of which they were not fully aware or had not considered important	India: citizens' report cards*
Providing objective, quantitative data that civil society can use to pressure agencies to improve services	India: citizens' report cards*
Demonstrating the economic and social benefits of village management of services, and alerting technical agencies to the need to incorporate vulnerable groups and address community conflicts caused by limited access to services	Indonesia: village water supply evaluation*
Providing political cover to government to take a politically sensitive decision to eliminate subsidies, thus mitigating negative consequences for influential "losers" from the policy change	Pakistan: Wheat-flour ration shop evaluation*
Developing methodologies to systematically document problems in the use of public funds that were widely suspected but that have not been possible to document	Uganda: Education Expenditure Tracking Surveys*

* Cases taken from Bamberger, Mackay and Ooi (2004, 2005).

** Cases taken from Bamberger and Kirk (editors) 2009.

key stakeholders, and what are their interests in the evaluation? Who are the main critics of the program, what are their concerns, and what would they like to happen? What kinds of evidence would they find most convincing? How can each of them influence the future direction of the program? What are the main concerns of different stakeholders with respect to the methodology? Are there sensitivities concerning the choice of quantitative or qualitative methods? How important are large sample surveys to the credibility of the evaluation?

Timing of the launch and completion of the evaluation. Many well-designed evaluations fail to achieve their intended influence because they were completed either too late (the critical decisions have already been made on future funding or program directions) or too early (before the questions being addressed are on the policy makers' radar screen).

Deciding what to evaluate. A successful evaluation will focus on a limited number of critical issues and hypotheses based on a clear understanding of the clients' information needs and of how the evaluation findings will be used. What do the clients need to know and what would they simply like to know? How will the evaluation findings be used? How precise and rigorous do the findings need to be?

Basing the evaluation on a program theory (logic) model. A program theory (logic) model developed in consultation with stakeholders is a good way to identify the key questions and hypotheses the evaluation should address. It is essential to ensure that clients and stakeholders and the evaluator share a common understanding with respect to the problem the program is addressing, what its objectives are, how they will be achieved, and what criteria the clients will use in assessing success.

Creating ownership of the evaluation. One of the key determinants of evaluation utilization is the extent to which clients and stakeholders are involved throughout the evaluation process. Do clients feel that they "own" the evaluation, or do they not know what the evaluation will produce until they receive the final report? The use of formative evaluation strategies that provide constant feedback to key stakeholders on how to use the initial evaluation findings to strengthen project implementation is also an effective way to enhance the sense of ownership. Communication strategies that keep clients informed and avoid their being presented with unexpected findings (that is, a "no surprises" approach) can create a positive attitude to the evaluation and enhance utilization.

Defining the appropriate evaluation methodology. A successful evaluation

must develop an approach that is both methodologically adequate to address the key questions and that is also understood and accepted by the client. Many clients have strong preferences for or against particular evaluation methodologies, and one of the factors contributing to underutilization of an evaluation may be client disagreement with, or lack of understanding of, the evaluation methodology.

*Process analysis and formative evaluation.*²⁵ Even when the primary objective of an evaluation is to assess program outcomes and impacts, it is important to “open-up the black box” to study the process of program implementation. Process analysis (the study of how the project is actually implemented) helps understand why certain expected outcomes have or have not been achieved; why certain groups may have benefited from the program and others have not; and to assess the *causes* of outcomes and impacts. Process analysis also provides a framework for assessing whether a program that has not achieved its objectives is fundamentally sound and should be continued or expanded (with certain modifications) or whether the program model has not worked—at least not in the contexts where it has been tried so far. Process analysis can suggest ways to improve the performance of an ongoing program, encouraging evaluation utilization as stakeholders can start to use these findings long before the final IE reports have been produced.

Evaluation capacity development is an essential tool to promote utilization because it not only builds skills, but it also promotes evaluation awareness.

Communicating the findings of the evaluation. Many evaluations have little impact because the findings are not communicated to potential users in a way that they find useful or comprehensible. The following are some guidelines for communicating evaluation findings to enhance utilization:

- Clarify what each user wants to know and the amount of detail required. Do users want a long report with tables and charts or simply a brief overview? Do they want details on each project location or a summary of the general findings?
- Understand how different users like to receive information. In a written report? In a group meeting with slide presentation? In an informal, personal briefing?
- Clarify if users want hard facts (statistics) or whether they prefer pho-

25. “An evaluation intended to furnish information for guiding program improvement is called a formative evaluation (Scriven 1991) because its purpose is to help form or shape the program to perform better” (Rossi, Lipsey, and Freeman, 2004, p. 34).

tos and narrative. Do they want a global overview, or do they want to understand how the program affects individual people and communities?

- Be prepared to use different communication strategies for different users.
- Pitch presentations at the right level of detail or technicality. Do not overwhelm managers with technical details, but do not insult professional audiences by implying that they could not understand the technicalities.
- Define the preferred medium for presenting the findings. A written report is not the only way to communicate findings. Other options include verbal presentations to groups, videos, photographs, meetings with program beneficiaries, and visits to program locations.
- Use the right language(s) for multilingual audiences.

Developing a follow-up action plan. Many evaluations present detailed recommendations but have little practical utility because the recommendations are never put into place—even though all groups might have expressed agreement. What is needed is an agreed action plan with specific, time-bound actions, clear definition of responsibility, and procedures for monitoring compliance.

9. Conclusions

ODA agencies are facing increasing demands to account for the effectiveness and impacts of the resources they have invested in development interventions. This has led to an increased interest in more systematic and rigorous evaluations of the outcomes and impacts of the projects, programs, and policies these agencies support. A number of high-profile and methodologically rigorous impact evaluations (IE) have been conducted in countries such as Mexico and Colombia, and many other countries are conducting IE of priority development programs and policies—usually with support from international development agencies.

Though many of these evaluations have contributed to improving the programs they have evaluated, much less progress has been made toward institutionalizing the processes of selection, design, implementation, dissemination, and use of IE. Consequently, the benefits of many of these evaluations have been limited to the specific programs they have studied, and the evaluations have not achieved their full potential as instruments for budget plan-

ning and development policy formulation. This chapter has examined some of the factors limiting the broader use of evaluation findings, and it has proposed guidelines for moving toward institutionalization of IE.

Progress toward institutionalization of IE in a given country can be assessed in terms of six dimensions: (a) Are the studies country-led and managed? (b) Is there strong buy-in from key stakeholders? (c) Have well-defined procedures and methodologies been developed? (d) Are the evaluations integrated into sector and national M&E systems? (e) Is IE integrated into national budget formulation and development planning? and (f) Are there policies and programs in place to develop evaluation capacity?

IE must be understood as only one of the many types of evaluation that are required at different stages of the project, program, or policy cycle, and it can only be effectively institutionalized as part of an integrated M&E system and not as a stand-alone initiative. A number of different IE designs are available, ranging from complex and rigorous experimental and quasi-experimental designs to less rigorous designs that are often the only option when working under budget, time, or data constraints or when the questions to be addressed do not merit the use of more rigorous and expensive designs.

Countries can move toward institutionalization of IE along one of at least three pathways. The first pathway begins with evaluations selected in an opportunistic or ad hoc manner and then gradually develops systems for selecting, implementing, and using the evaluations (for example, the SINERGIA M&E system in Colombia). The second pathway develops IE methodologies and approaches in a particular sector that lay the groundwork for a national system (for example, Mexico); the third path establishes IE from the beginning as a national system although it may be refined over a period of years or even decades (Chile). The actions required to institutionalize the IE system were discussed. It is emphasized that IE can only be successfully institutionalized as part of an integrated M&E system and that efforts to develop a stand-alone IE system are ill advised and likely to fail.

Conducting a number of rigorous IEs in a particular country does not guarantee that ministries and agencies will automatically increase their demand for more. In fact, a concerted strategy has to be developed for creating demand for IE as well as other types of evaluation. Though it is essential to strengthen the *supply* of evaluation specialists and agencies able to implement evaluations, experience suggests that creating the *demand* for evaluations is equally if not more important. Generating demand requires a combination of incentives (carrots), sanctions (sticks), and positive messages from

key figures (sermons). A key element of success is that IE be seen as an important policy and management tool in one or more of the following areas: providing guidance on resource allocation, helping ministries in their policy formulation and analytical work, aiding management and delivery of government services, and underpinning accountability.

Evaluation capacity development (ECD) is a critical component of IE institutionalization. It is essential to target five different stakeholder groups: agencies that commission, fund, and disseminate IEs; evaluation practitioners; evaluation users; groups affected by the programs being evaluated; and public opinion. Although some groups require the capacity to design and implement IE, others need to understand when an evaluation is needed and how to commission and manage it. Still others must know how to disseminate and use the evaluation findings. An ECD strategy must give equal weight to all five groups and not, as often happens, focus mainly on the researchers and consultants who will conduct the evaluations.

Many IEs rely mainly on the generation of new survey data, but there are often extensive secondary data sources that can also be used. Although secondary data have the advantage of being much cheaper to use and can also reconstruct baseline data when the evaluation is not commissioned until late in the project or program cycle, they usually have the disadvantage of not being project specific. A valuable, but frequently ignored source of evaluation data is the monitoring and administrative records of the program or agency being evaluated. The value of these data sources for the evaluation can be greatly enhanced if the evaluators are able to coordinate with program management to ensure monitoring and other data are collected and organized in the format required for the evaluation. Where possible, the evaluation should use a mixed-method design combining quantitative and qualitative data-collection and analysis methods. This enables the evaluation to combine the breadth and generalizability of quantitative methods with the depth provided by qualitative methods.

Many well-designed and potentially valuable evaluations (including IEs) are underutilized for a number of reasons, including lack of ownership by stakeholders, bad timing, failure to address client information needs, lack of follow-up on agreed actions, and poor communication and dissemination. An aggressive strategy to promote utilization is an essential component of IE institutionalization.

The roles and responsibilities of ODA agencies and developing country governments in strengthening the institutionalization of IE

ODA agencies, who provide major financial and technical support for IE, will continue to play a major role in promoting institutionalization of IE. One major responsibility must be an active commitment to move from *ad hoc* support to individual IEs that are of particular interest to the donor country, to a genuine commitment to helping countries develop an IE system that serves the interests of national policymakers and line ministries. This requires a full commitment to the 2005 Paris Declaration on Aid Effectiveness, in particular to support for country-led evaluation. It also requires a more substantial and systematic commitment to evaluation capacity development, and a recognition of the need to develop and support a wider range of IE methodologies designed to respond more directly to country needs and less on seeking to impose methodologically rigorous evaluation designs that are often of more interest to ODA research institutions than to developing country governments. This emphasizes the importance of recognizing the distinction between the *technical* and the *substantive* definitions of IE, and accepting that most country IE strategies should encompass both definitions in order to draw on a broader range of methodologies to address a wide range of policy and operational questions.

For their part, developing country governments must invest the necessary time and resources to ensure they fully understand the potential benefits and limitations of IE and the alternative methodologies that can be used. Usually one or more ODA agencies will be willing to assist governments wishing to acquire this understanding, but governments must seek their own independent advisors so that they do not depend exclusively on the advice of a particular donor who may have an interest in promoting only certain types of IE methodologies. Some of the sources of advice include: national and international consultants, national and regional evaluation conferences and networks (see the UNICEF publication by Segone and Ocampo 2006), a large number of easily accessible websites²⁶ and study-tours to visit countries that have made progress in institutionalizing their IE systems. The key steps and strategic options discussed in this paper, together with the references to the literature, should provide some initial guidelines on how to start the institutionalization process.

26. The American Evaluation Association website (eval.org) has one of the most extensive reference lists on evaluation organizations, but many other agencies have similar websites.

References

- Bamberger, M. (2009) Strengthening the evaluation of program effectiveness through reconstructing baseline data. *Journal of Development Effectiveness*, 1(1), (March 2009).
- Bamberger, M. and Kirk, A. (eds.) (2009) *Making smart policy: Using impact evaluation for policymaking*. Thematic group for Poverty Analysis, Monitoring and Impact Evaluation, World Bank.
- Bamberger, M. (2009) *Institutionalizing Impact Evaluation within the Framework of a Monitoring and Evaluation System*. Washington, D.C.: Independent Evaluation Group, World Bank.
- Bamberger, M. and Fujita, N. (2008) *Impact Evaluation of Development Assistance*. FASID Evaluation Handbook. Tokyo: Foundation for Advanced Studies for International Development.
- Bamberger, M. (2006a) *Conducting Quality Impact Evaluations under Budget, Time, and Data Constraints*. Washington, D.C.: World Bank.
- Bamberger, M. (2006b) "Evaluation Capacity Building." In *Creating and Developing Evaluation Organizations: Lessons learned from Africa, Americas, Asia, Australasia and Europe*, ed. M. Segone. Geneva: UNICEF.
- Bamberger, M., J. Rugh, and L. Mabry. (2006) *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Thousand Oaks, CA: Sage Publications.
- Bamberger, M., K. Mackay, and E. Ooi. (2005) *Influential Evaluations: Detailed Case Studies*. Washington, D.C.: Operations Evaluation Department, World Bank.
- Bamberger, M., K. Mackay, and E. Ooi. (2004) *Influential Evaluations: Evaluations that Improved Performance and Impacts of Development Programs*. Washington, D.C.: Operations Evaluation Department, World Bank.
- DFID (Department for International Development, UK). (2005) Guidance on Evaluation and Review for DFID Staff. <http://www.dfid.gov.uk/aboutdfid/performance/files/guidance-evaluation.pdf>.
- Kane, M. and Trochim, W. (2007) *Concept mapping for planning and evaluation*. Thousand Oaks, CA: Sage Publications.
- Mackay, K. (2007) *How to Build M&E Systems to Support Better Government*. Washington, D.C.: Independent Evaluation Group, World Bank.
- Mackay, K., Lopez-Acevedo, G., Rojas, F. & Coudouel, A. (2007) *A diagnosis of Colombia's National M&E system: SINERGIA*. Independent Evaluation

- Group, World Bank.
- OECD-DAC (Organization of Economic Co-operation and Development, Development Advisory Committee). (2002) *Glossary of Key Terms in Evaluation and Results Based Management*. Paris: OECD.
- Patton, M.Q. (2008) *Utilization-Focused Evaluation* (4th ed.). Thousand Oaks, CA: Sage Publications.
- Picciotto, R. (2002) "International Trends and Development Evaluation: The Need for Ideas." *American Journal of Evaluation*, 24 (2): 227–34.
- Ravallion, M. (2008) Evaluation in the Service of Development. Policy Research Working Paper, No. 4547. Washington, D.C.: World Bank.
- Rossi, P., M. Lipsey, and H. Freeman. (2004) *Evaluation: a Systematic Approach* (7th ed.). Thousand Oaks, CA: Sage Publications.
- Schiavo-Campo, S. (2005) *Building country capacity for monitoring and evaluation in the public sector: selected lessons of international experience*. Washington, D.C.: Operations Evaluation Department, World Bank.
- Scriven, M. (2009) "Demythologizing causation and evidence." In Donadlson, S., Christie, C., and Mark, M. *What counts as credible evidence in applied research and evaluation practice?* Thousand Oaks, CA: Sage Publications.
- Scriven, M. (1991) *Evaluation Thesaurus* (4th ed.). Newbury Park, CA: Sage Publications.
- Segone and Ocampo. (eds.) (2006) *Creating and Developing Evaluation Organizations: Lessons from Africa, Americas, Asia, Australasia and Europe*. Geneva: UNICEF.
- Segone. (ed.) (2008) *Bridging the gap: the role of monitoring and evaluation in evidence-based policymaking*. Geneva: UNICEF.
- Segone. (ed.) (2009) *Country-led monitoring and evaluation systems: Better evidence, better policies, better development results*. Geneva: UNICEF.
- White, H. (2006) Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank. Working Paper, No. 38268. Washington, D.C.: World Bank.
- Wholey, J.S., J. Scanlon, H. Duffy, J. Fukumoto, and J. Vogt. (1970) *Federal Evaluation Policy: Analyzing the Effects of Public Programs*. Washington, D.C.: Urban Institute.
- World Bank, OED (Operations Evaluation Department). (2004) *Monitoring and Evaluation: Some Tools, Methods and Approaches*. Washington, D.C.: World Bank.
- Zaltsman, A. (2006) *Experience with institutionalizing monitoring and evaluation systems in five Latin American countries: Argentina, Chile, Colombia,*

Costa Rica and Uruguay. Washington, D.C.: Independent Evaluation Group, World Bank.

Editor's Notes

In Trends in Development Assistance Series 5, we surveyed the current status of Japan's development aid evaluation, sorted out various public policy evaluation schemes including the ODA evaluation, and then described examples of Japan's assistance for evaluation capacity development and its support initiatives for institutionalizing evaluation systems. With regard to evaluation capacity development and institutionalization of evaluation systems, the interest in impact evaluations has been rising in recent years. On that note, we also reported on The World Bank's assistance for impact evaluations in Latin America.

In Chapter 1 ("Development Assistance Evaluation in Japan: Challenges and Outlook"), Muta and Minamoto provide a comprehensive analysis of the history of ODA evaluations in Japan, the development of evaluation systems, basic policies and current status. Based on that analysis, they state that, with regard to program- and policy-level evaluations, structural evaluations including higher-level subjects will become possible when the program approach is realized as intended with the birth of the new JICA.

Interestingly, they suggest including third parties in mid-term evaluations of JICA's technical cooperation, which so far have been conducted as internal evaluations. They also emphasize the importance of participatory evaluation which involves stakeholders on the side of recipient countries, whereas ODA evaluations in the past tended to be conducted mostly by donor countries. They also suggest that, rather than evaluating all projects, the management of individual projects should be strengthened by monitoring and emphasis should be placed on higher-level, impact-related evaluations. In other words, they are suggesting a separation of monitoring and evaluation, or at least a differentiation of roles between performance measurement and evaluation. Now that evaluation has taken firm roots in ODA, the roles of evaluation should be examined anew in order to prevent it from becoming a "mere formality." The risk of evaluation becoming a ritualized process is a shared concern of the authors.

Evaluations of major aid schemes - grant aid cooperation, technical cooperation and loan aid - used to be implemented by the Ministry of Foreign Affairs, JICA and JBIC, respectively. The merger of JICA and JBIC will provide a great opportunity for building a new evaluation system, and we have great hopes for new developments in the near future.

In Chapter 2 (“ODA Evaluation and Policy Evaluation: Status of Accountability and Transparency in Japan”), Yamaya clarifies terminological and conceptual confusions in various fields in Japan where evaluations and similar activities are conducted, and where the ODA evaluation stands in that scheme.

Yamaya compares the ODA evaluation and policy evaluation and concludes that they are different in terms of the contents of and procedures for securing accountability; the degrees of integration into budgeting and human resource management; and the complexity of systems. He also points out that there is a difficulty in the ODA evaluation in that if policy-level subjects are included in its scope, one must consider even contributions to Japan's diplomatic policies. On the other hand, the domestic policy evaluation has a different difficulty in that there are fewer alternative policy instruments and interests are more complex than in the area of ODA. Yamaya concludes that, in any case, both the domestic policy evaluation and ODA evaluation must eventually converge on the administrative management-type evaluation. However, such administrative management-type evaluation is only used as a management tool. For this reason, the author emphasizes the importance of developing evaluators who recognize the diversity of evaluation and can respond to other types of requirements.

In Chapter 3 (“Evaluation Capacity Development: A Practical Approach to Assistance”), Haraguchi and Miyazaki draw from their experience as consultants in Indonesia and Vietnam to describe in detail cases of assistance in evaluation capacity development.

In countries where the concept of evaluation is not yet widely accepted, it is necessary to stimulate the demand for evaluation and not to leave evaluations entirely up to donors. The authors state that, as with the cooperation of (former) JBIC, the methods of introducing the international standards of evaluation techniques, and conducting joint donor and recipient evaluations thereby increasing the recipient country's interest in evaluation, were very effective.

Haraguchi and Miyazaki also suggest that, the keys to successful cooperation for evaluation capacity development are providing institutional assistance, increasing the awareness of evaluation, spreading the knowledge of not only ex-post evaluations but also mid-term evaluations and monitoring, effective use of seminars, and utilizing and integrating outsourced evaluations into evaluation management. However, they also state that, when local

consultants are utilized in an outsourced evaluation, the government agency in charge needs to be capable of managing the external evaluation and controlling the quality of its findings.

In Chapter 4 (“Evaluation Capacity Development in the Asia-Pacific Region: A Proposal for an Asia-Pacific Evaluation Association Network (APEA NET)”), Hirono describes issues related to evaluation capacities in the region (including institutional and organizational ones) and proposes, as a means to develop such capacities, to establish a regional network.

In Cambodia, UNDP, New Zealand and the United Kingdom provide assistance for evaluation capacity development at the country level. Mongolia is receiving similar assistance from UNDP. In the Asia-Pacific region, Hirono states that India, Malaysia, the Philippines, Singapore and Sri Lanka are showing rapid progress in monitoring and evaluation programs, although there are various differences among countries due to varying political systems and degrees of decentralization.

Academic associations on evaluation are increasing in numbers. Japan, Australia and South Korea have already established such associations. Bangladesh, India, Malaysia, Pakistan, Nepal and Sri Lanka. Thailand and Vietnam are in the process of establishing associations, and Cambodia, Indonesia, Laos, Mongolia and the Philippines are expected to follow suit. On the other hand, Myanmar shows no such indication, and China does not have an evaluation society but it does have an established and strong evaluation mechanism. In light of these circumstances, Hirono proposes establishing an evaluation network in the Asia-Pacific region. The network would support the establishment of national academic societies, and their activities which contribute to more efficient and effective development of economies and societies by promoting the culture of evaluation, and training evaluation specialists.

In Chapter 5 (“Institutionalizing Impact Evaluation Systems in Developing Countries: Challenges and Opportunities for ODA Agencies”), Bamberger introduces cases in Latin America, Africa and China with regard to the current positioning of the impact evaluation in the institutionalization of monitoring and evaluation.

The necessity for evaluating impacts of development assistance has been emphasized in recent years. The term “impact evaluation” sometimes refers to an evaluation of long-term effects, which is the OECD-DAC’s definition.

(In this case, evaluation methods are of no concern.) At other times, it refers to specific technical and statistical methods. (In this case, methods are limited to those using counterfactuals, but the time horizon of effects is of no concern.) This chapter describes the impact evaluation in a broader sense that includes both.

Bamberger first introduces cases in Latin America (specifically Mexico, Colombia and Chile), describing the process of moving from individual impact evaluation studies to institutionalization of impact evaluations. Each country took a different path. Colombia started with ad hoc implementation of strict impact evaluations with assistance from The World Bank, but the involvement of the Ministry of Planning opened the door to a wider variety of methods, and eventually, technical criteria and procedures were stipulated in policy documents. In the case of Mexico, it started as an evaluation of the Progresa Conditional Cash Transfer programs designed for a priority objective of poverty reduction, and it led to the establishment of a country-level evaluation system. Similar cases are reported in Uganda and China as well. Chile is an example where the impact evaluation was introduced as part of a monitoring and evaluation system for the entire government. Chile's system is unique in that its objective is limited: only the Ministry of Finance uses it for budgeting purposes.

The World Bank is assisting 20 African countries implement impact evaluations and develop related capacities. Lessons learned by The World Bank from its evaluation capacity development (ECD) assistance are instructive. For example, the Bank states that effective capacity development should target five stakeholder groups: organizations that commission evaluations, organizations/individuals who conduct them, those who use them, those who are affected by them, and the general public including academic and civil societies. We suspect that normal ECD activities include only the first two groups in their scope.

For impact evaluations, the author recommends the mixed methods approach which take into account the strengths and weaknesses of both qualitative and quantitative methods. Related to this, Bamberger states that governments of developing countries should not rely solely on the advice of a single donor who is only interested in a certain method of impact evaluation, and instead seek independent advisors (consultants and various evaluation-related networks), because, as described above, the term "impact evaluation" is used to mean two different things, and both should be covered by a wide variety of methods. On the donors' side, the author argues that it is important

for them to shift away from assisting individual impact evaluations to supporting the construction of evaluation systems that will be useful to politicians and related ministries of partner countries.

What we can see from these five chapters is that, while ongoing cooperation for establishing evaluation systems and developing evaluation capacity, evaluation has a unique characteristic in that once a system is established it is immediately faced with the risk of being reduced to a mere formality. Evaluation is not a type of endeavor which will go smoothly once the system and methods are established. There are no two evaluations exactly alike. As evaluations become more accepted and evaluation systems established, it becomes more important to constantly remind ourselves for what and how they should be utilized. To avoid “having the form but not the spirit,” the meaning of evaluation should be examined in each case, and that is where evaluation specialists have a role to play.

March 2009

Nobuko FUJITA
Editor

Authors

Chapter 1 Development Assistance Evaluation in Japan: Challenges and Outlook

MUTA, Hiromitsu

Executive Vice President for Finance, Tokyo Institute of Technology
Professor, Graduate School of Decision Science and Technology,
Tokyo Institute of Technology

MINAMOTO, Yuriko

Associate Professor, Graduate School of Governance Studies, Meiji
University

Chapter 2 ODA Evaluation and Policy Evaluation: Status of Accountability and Transparency in Japan

YAMAYA, Kiyoshi

Professor, Faculty of Policy Studies, Doshisha University

Chapter 3 Evaluation Capacity Development: A Practical Approach to Assistance

HARAGUCHI, Takako

Senior Consultant, International Development Associates Ltd.

MIYAZAKI, Keishi

Deputy General Manager, Planning Department, OPMAC
Corporation

Chapter 4 Evaluation Capacity Development in the Asia-Pacific Region: A Proposal for an Asia-Pacific Evaluation Association Network (APEA NET)

HIRONO, Ryokichi

Professor Emeritus, Seikei University

Chair, International Affairs Department, Japan Evaluation Society

Chapter 5 Institutionalizing Impact Evaluation Systems in Developing Countries: Challenges and Opportunities for ODA Agencies

BAMBERGER, Michael

Social Development and Program Evaluation Consultant
(Former World Bank Senior Sociologist)

Editors

MINATO, Naonobu

Acting Director, International Development Research Institute
(IDRI), Foundation for Advanced Studies on International
Development (FASID)

FUJITA, Nobuko

Deputy Director, IDRI, FASID

FASiD